

# nature

## LOST IN THE CITY

Grid-planned urban environments hamper future spatial navigation skills

### Coronavirus

Why the WHO took so long to acknowledge airborne transmission

### Financial risks

Priorities for the regulation of cryptocurrencies

### Copy errors

The destabilizing effects of genome doubling in cancer

# Nature.2022.04.09

[Sat, 09 Apr 2022]

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

# This Week

- **[The EU can simultaneously end dependence on Russia and meet climate goals](#)** [ 05 April 2022]  
Editorial • Nations that relinquish reliance on Russia's fossil fuels should quickly transition to clean energy.
- **[Time to recognize authorship of open data](#)** [ 06 April 2022]  
Editorial • The open data revolution won't happen unless the research system values the sharing of data as much as authorship on papers.
- **[Give refugees dignity, wherever they are](#)** [ 05 April 2022]  
World View • Displaced children and adults can wait decades to return home or resettle — research and policy must catch up.
- **[Global cities are sinking — and humans are partly to blame](#)** [ 31 March 2022]  
Research Highlight • Some coastal cities are subsiding by dozens of millimetres per year, making them even more vulnerable to sea-level rise triggered by climate change.
- **[A global sand crisis looms — here's how to dodge it](#)** [ 28 March 2022]  
Research Highlight • The right strategies could help to reduce sand mining and the associated environmental destruction.
- **[‘Squashed’ atomic nucleus sets speed record](#)** [ 30 March 2022]  
Research Highlight • A form of the element lutetium exists for only a few hundred nanoseconds before decaying.
- **[A blast of light gives Styrofoam and other plastics new purpose](#)** [ 31 March 2022]  
Research Highlight • Polystyrene products can be recycled into a widely used compound thanks to a light-driven reaction.
- **[Secret of the birds: nearly 300 species fool foes the same way](#)** [ 29 March 2022]  
Research Highlight • The ‘broken wing’ trick has evolved in more than 50 bird families, hinting at its success in protecting eggs and chicks.

- **Marauding crazy ants come to grief when a fungus comes to call** [ 28 March 2022]

Research Highlight • Insect's populations defy many conventional pesticides but collapse after infection by a microbial pathogen.

- **Bacteria in China's livestock amass weapons against antibiotics** [ 31 March 2022]

Research Highlight • Escherichia coli sampled from pigs and other farm animals show a steep rise in antibiotic-resistance genes.

- EDITORIAL
- 05 April 2022

# The EU can simultaneously end dependence on Russia and meet climate goals

Nations that relinquish reliance on Russia's fossil fuels should quickly transition to clean energy.



The Nord Stream 2 pipeline was designed to boost Europe's Russian gas imports. Credit: Andrey Rudakov/Bloomberg/Getty

Vladimir Putin's invasion of Ukraine initiated Europe's largest war since the Second World War. The United States and many European nations have responded to the Russian president's actions by cutting trading, financial and scientific links with Russia on a scale never previously seen for a permanent member of the United Nations Security Council.

But [Europe's energy needs are stopping many countries](#) from taking even stronger action on the economic front. Russia supplies the European Union with around 40% of its natural gas, as well as about 25% of its oil and almost 50% of its coal. And this trade is continuing. It's a weakness that Putin has exploited. Last week, he passed a decree stating that the gas supplies of "unfriendly" nations would be turned off if customers didn't pay in roubles (rather than US dollars or euros).

Putin's move seems to be intended as both a retaliation against sanctions and an attempt to shore up the rouble; the currency lost around half of its value in the 2 weeks after the 24 February invasion, reaching a record low of roughly 150 roubles to the dollar. It has since rebounded, but the outlook remains volatile.

European leaders rightly rejected Putin's demands and, for now at least, the Kremlin looks to be backing down. Researchers and analysts think Putin is unlikely to cut off gas supplies because Russia would lose a large fraction of its daily €700-million (US\$771-million) income from the EU — money that it is currently using to support its military. And Russia is not in a position to recoup this loss by redirecting its gas supplies to friendlier countries, because this would necessitate the construction of new infrastructure, which cannot be done quickly.

Whatever happens, the threat is a sign that the EU needs to accelerate its efforts to relinquish its dependence on Russia's fossil fuels. It also underscores something that researchers who study climate, energy and economics have been saying for decades: that climate security and energy security are linked.

Researchers contacted by *Nature* say that European countries should be able to get through the next winter without Russian imports or power outages. However, to do so would require immediate action on many fronts,

including intensive international cooperation to boost imports of natural gas from other countries; the launch of a burst of clean-energy projects; and the introduction of a host of energy-conservation and efficiency measures, possibly including energy rationing.

If the immediate goal is to keep the lights on, the long-term goal must be decarbonization, which will allow Europe to simultaneously end its dependence on Russia and meet its climate goals. As this week's report from the Intergovernmental Panel on Climate Change indicates (see [go.nature.com/3k7vgu0](https://go.nature.com/3k7vgu0)), it means rapidly replacing the fossil fuels that power national and regional electricity systems with renewable alternatives. It also means deploying vehicles that run on electricity or renewably sourced hydrogen, and retrofitting homes and businesses to use less — and produce more — energy. Moreover, a suite of energy and carbon-capture strategies will be needed to clean up heavy industry. None of this will be easy, but if the world wants both a more secure and a cleaner future, we don't have a choice.

Scientists and environmentalists have been making this case for decades, warning governments that fossil fuels destabilize not only the climate, but also geopolitics, by creating dependencies on problematic regimes.

Whether or not European countries decide to stop buying Russian gas, they will almost certainly experience considerable economic pain as prices continue to increase. With many businesses unable to withstand the coming shocks unaided, and the resulting potential for job losses, governments will have no option but to step in with relief.

European leaders are acutely aware that they are financing the enemy at their gates. They must remain united, and coordinate and accelerate the clean-energy transition — action that will be required if they are to achieve the goal set out in the Paris climate agreement of limiting global warming to 1.5 °C above pre-industrial temperatures.

In the short term, the need for energy security will probably see more power than usual generated using fossil fuels, but the overall message cannot now be faulted: European leaders must understand that decarbonization is the answer to both energy and climate security. And if they manage to lay the

groundwork for a cleaner future as part of their response to the war in Ukraine, theirs will be a lesson for the world.

*Nature* **604**, 7-8 (2022)

*doi:* <https://doi.org/10.1038/d41586-022-00920-y>

---

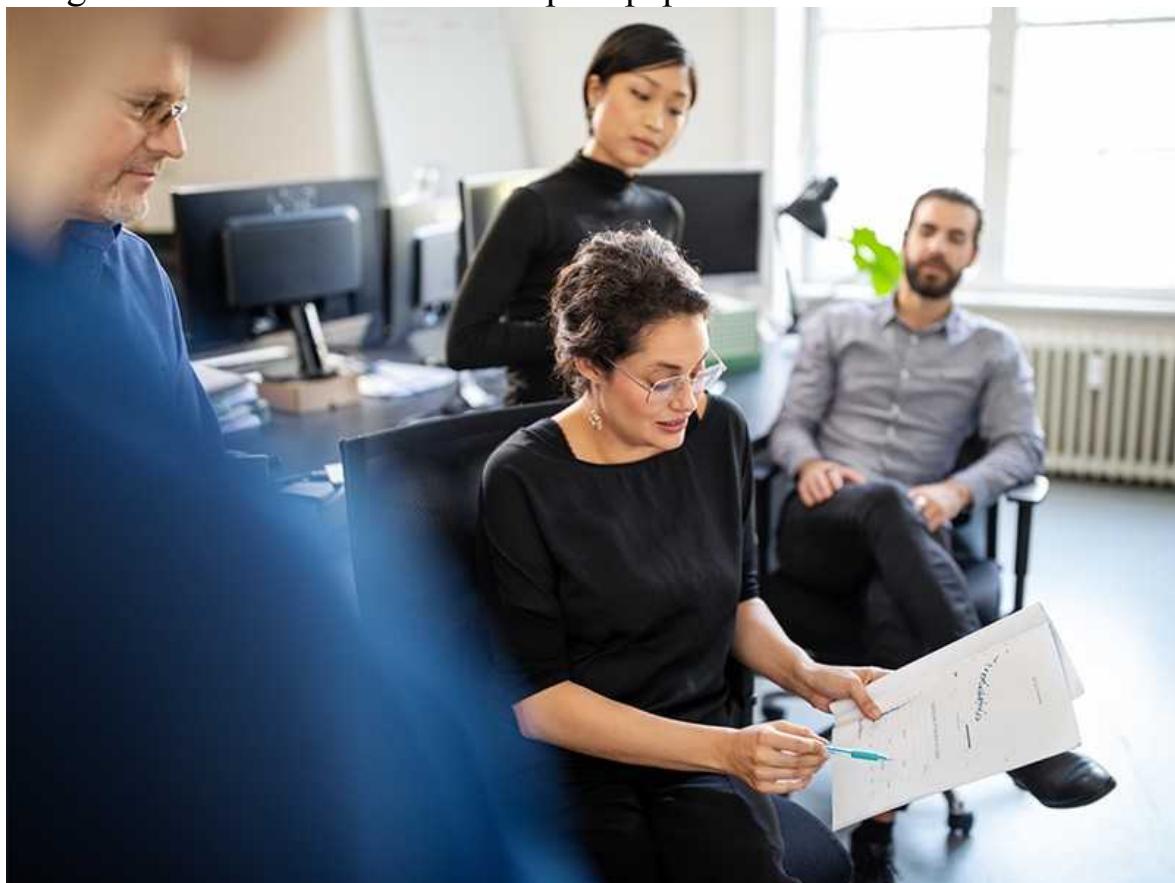
This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-022-00920-y>.

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

- EDITORIAL
- 06 April 2022

# Time to recognize authorship of open data

The open data revolution won't happen unless the research system values the sharing of data as much as authorship on papers.



Researchers willing to share data are not always appropriately credited by those evaluating research. Credit: Getty

At times, it seems there's an unstoppable momentum towards the principle that data sets should be made widely available for research purposes (also

called open data). Research funders all over the world are [endorsing the open data-management standards](#) known as the FAIR principles (which ensure data are findable, accessible, interoperable and reusable). Journals are increasingly asking authors to make the underlying data behind papers accessible to their peers. Data sets are accompanied by a digital object identifier (DOI) so they can be easily found. And this citability helps researchers to get credit for the data they generate.

But reality sometimes tells a different story. The world's systems for evaluating science do not (yet) value openly shared data in the same way that they value outputs such as journal articles or books. Funders and research leaders who design these systems accept that there are many kinds of scientific output, but many reject the idea that there is a hierarchy among them.

In practice, those in powerful positions in science tend not to regard open data sets in the same way as publications when it comes to making hiring and promotion decisions or awarding memberships to important committees, or in national evaluation systems. The open-data revolution will stall unless this changes.

This week, Richard Bethlehem at the University of Cambridge, UK, and Jakob Seidlitz at the University of Pennsylvania in Philadelphia and their colleagues publish research describing brain development 'charts' ([R. A. I. Bethlehem et al. Nature https://doi.org/10.1038/s41586-022-04554-y; 2022](#)). These are analogous to the charts that record height and weight over the course of a person's life, which researchers and clinicians can access.

This work has never been done on such a scale: typically in neuroscience, studies are based on relatively small data sets. To create a more globally representative sample, the researchers aggregated some 120,000 magnetic resonance imaging scans from more than 100 studies. Not all the data sets were originally available for the researchers to use. In some cases, for example, formal data-access agreements constrained how data could be shared.

Some of the scientists whose data were originally proprietary became active co-authors on the paper. By contrast, researchers whose data were accessible

from the start are credited in the paper's citations and acknowledgements, as is the convention in publishing.

Such a practice is neither new nor confined to a specific field. But the result tends to be the same: that authors of openly shared data sets are at risk of not being given credit in a way that counts towards promotion or tenure, whereas those who are named as authors on the publication are more likely to reap benefits that advance their careers.

Such a situation is understandable as long as authorship on a publication is the main way of getting credit for a scientific contribution. But if open data were formally recognized in the same way as research articles in evaluation, hiring and promotion processes, research groups would lose at least one incentive for keeping their data sets closed.

Universities, research groups, funding agencies and publishers should, together, start to consider how they could better recognize open data in their evaluation systems. They need to ask: how can those who have gone the extra mile on open data be credited appropriately?

There will always be instances in which researchers cannot be given access to human data. Data from infants, for example, are highly sensitive and need to pass stringent privacy and other tests. Moreover, making data sets accessible takes time and funding that researchers don't always have. And researchers in low- and middle-income countries have concerns that their data could be used by researchers or businesses in high-income countries in ways that they have not consented to.

But crediting all those who contribute their knowledge to a research output is a cornerstone of science. The prevailing convention — whereby those who make their data open for researchers to use make do with acknowledgement and a citation — needs a rethink. As long as authorship on a paper is significantly more valued than data generation, this will disincentivize making data sets open. The sooner we change this, the better.

*Nature* **604**, 8 (2022)

doi: <https://doi.org/10.1038/d41586-022-00921-x>

---

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-022-00921-x>

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

- WORLD VIEW
- 05 April 2022

# Give refugees dignity, wherever they are



Displaced children and adults can wait decades to return home or resettle — research and policy must catch up.

- [Serena Parekh](#) 9

More than four million people have fled Ukraine since Vladimir Putin's brutal invasion. Half the nation's children are displaced. It is crucial to study states' moral obligations around asylum and resettlement and to consider how these outcomes affect individual well-being, public health, economics and political stability. Yet in focusing on these end points too narrowly, researchers and policymakers can overlook another set of extremely important questions.

What happens to someone who has become a refugee while they wait to join a new country or to return home? On average, a person remains a refugee for more than a decade — 25 years for those fleeing war. Fewer than 1% of the world's 26.6 million refugees are resettled each year. In their book *Refuge* (2017), Alexander Betts and Paul Collier estimated that the world spends around US\$75 billion on the 10% who seek asylum in wealthy countries, and just \$5 billion on the remaining 90%. In this liminal state, the problems of how refugees live, how they are treated by host countries and their citizens, and whether they have access to basic human rights are, in my experience, insufficiently addressed by scholars, policymakers and concerned members of the public.

The solutions found for Ukrainians bear scrutiny. What was unthinkable two months ago — that Europe could feed, house and economically integrate a huge number of refugees quickly and with little animosity — has become a reality. What happens next and how such creativity might be replicated elsewhere are worthy of attention.

As I showed in my second book on the ethics of the global refugee crisis, *No Refuge* (2020), most displaced people face three options: camps; urban centres; or dangerous journeys in search of permanent residency or citizenship. None of these gives individuals a minimum level of dignity.

Most refugee camps are meant to be temporary. Consequently, when they become long-term structures, they often lack adequate food, health care, meaningful work or security (notably against gender-based violence). The Dadaab site in Kenya, for example, was built in 1991 to shelter 90,000 people fleeing civil war in Somalia, but housed half a million at its peak. Three decades on, generations have been born and raised there. In 2015, food rations were 30% less than the minimum recommended by the United Nations.

More than 60% of refugees thus prefer to live informally in towns and cities. Here, fewer than one in ten have access to aid. They live precariously, without schooling, health care or routes to citizenship.

Some try to reach Europe, the United States or Australia to apply for asylum. These regions have brutal deterrence policies. More than 20,000

migrants have died crossing the Mediterranean from North Africa and Turkey to Italy, Spain and Greece since 2014. People escaping war in Syria and Afghanistan, including women and children, have been beaten back from Europe with water cannons, tear gas, guard dogs and razor-wire fences.

Even after refugees apply for asylum, they often have to live in degrading conditions. The Moria refugee camp in Greece, Europe's largest until it burnt down in 2020, had been likened to an open-air prison: overcrowded, with appalling sanitary conditions, rampant infectious diseases, and violence. Children as young as ten had attempted suicide.

The suffering of the Ukrainian people is immense. The response — compassion, solidarity and bureaucratic efficiency — should serve as a model for how the world accommodates all refugees. Within the first week of the war, Poland and other neighbours of Ukraine ran free trains and built reception centres; locals greeted arrivals with bowls of borscht and teddy bears; systems were set up to connect refugees with citizens willing to house and feed them; and online job boards matched Ukrainians seeking work with employers across Europe.

The most heartening development came less than two weeks after Russia invaded: the European Union activated the Temporary Protection Directive. This protocol gives Ukrainian refugees access to important social goods without the long process of applying for asylum. They receive the ability to live and move freely in the EU, and the rights to work, education and health care. The status is granted for one year, renewable for up to three.

These rights are precisely what traumatized Ukrainians need and deserve, and precisely what the vast majority of the world's refugees lack. Such provisions are not sufficient for thriving. But they are necessary for dignity.

In short, the innovations giving assistance to displaced Ukrainians must continue, and must spread beyond cultural-affinity groups, to non-European refugees in Europe and elsewhere who have experienced similar trauma and violence. Nations' duties under international agreements extend beyond asylum, to the 90%: those living 'temporarily' in urban centres and camps.

These people, too, need quality health care, education and ways to work and move freely.

Politicians and citizens have a moral obligation to provide all refugees with the basic conditions of dignity, without discrimination — no matter what their national origin or religious background.

*Nature* **604**, 9 (2022)

*doi:* <https://doi.org/10.1038/d41586-022-00938-2>

---

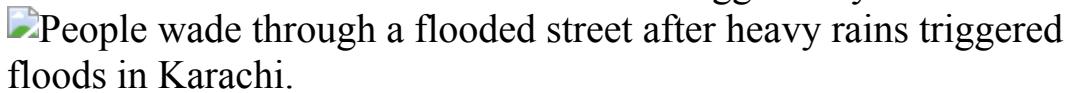
This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-022-00938-2>

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

- RESEARCH HIGHLIGHT
- 31 March 2022

# Global cities are sinking — and humans are partly to blame

Some coastal cities are subsiding by dozens of millimetres per year, making them even more vulnerable to sea-level rise triggered by climate change.



People slosh down a flooded street in the port city of Karachi, Pakistan, which is sinking by more than 10 millimetres per year. Credit: Asif Hassan/AFP via Getty

Many coastal cities around the world are sinking as sea levels rise, a combination of trends that drastically increases the risk of urban flooding<sup>1</sup>.

## Access options

Subscribe to Nature+

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

\$29.99

monthly

[Subscribe](#)

Subscribe to Journal

Get full journal access for 1 year

\$199.00

only \$3.90 per issue

[Subscribe](#)

All prices are NET prices.

VAT will be added later in the checkout.

Tax calculation will be finalised during checkout.

Buy article

Get time limited or full article access on ReadCube.

\$32.00

[Buy](#)

All prices are NET prices.

### **Additional access options:**

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

*Nature* **604**, 10 (2022)

*doi:* <https://doi.org/10.1038/d41586-022-00911-z>

## **References**

1. Wu, P.-C., Wei, M. (M.) & D'Hondt, S. *Geophys. Res. Lett.* <https://doi.org/10.1029/2022GL098477> (2022).

---

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-022-00911-z>

- RESEARCH HIGHLIGHT
- 28 March 2022

# A global sand crisis looms — here's how to dodge it

The right strategies could help to reduce sand mining and the associated environmental destruction.

 Laborers and machines work inside the Kuakhai river digging sand to transport to the area construction sites in India.

Sand for construction is collected from a river bed in India, where estimated demand for the material will more than double by 2060. Credit: NurPhoto via Getty

Global use of sand for building will surge by 45% by 2060, even if population and economic growth are moderate, according to new modelling<sup>1</sup>. But the analysis suggests that there's hope: if countries cooperate, building-sand demand in 2060 could be 30% of projected levels.

## Access options

Subscribe to Nature+

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

\$29.99

monthly

[Subscribe](#)

Subscribe to Journal

Get full journal access for 1 year

\$199.00

only \$3.90 per issue

[Subscribe](#)

All prices are NET prices.

VAT will be added later in the checkout.

Tax calculation will be finalised during checkout.

Buy article

Get time limited or full article access on ReadCube.

\$32.00

[Buy](#)

All prices are NET prices.

### **Additional access options:**

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

*Nature* **604**, 10 (2022)

*doi:* <https://doi.org/10.1038/d41586-022-00876-z>

## **References**

1. Zhong, X. Deetman, S. Tukker, A. & Behrens, P. *Nature Sust.* [\(2022\).](https://doi.org/10.1038/s41893-022-00857-0)

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

- RESEARCH HIGHLIGHT
- 30 March 2022

## ‘Squashed’ atomic nucleus sets speed record

A form of the element lutetium exists for only a few hundred nanoseconds before decaying.

 Artist's impression of the decay of a  $^{149}\text{Lu}$  nucleus into a  $^{148}\text{Yb}$  nucleus and a proton.

A nucleus (green and yellow) of the element lutetium emits a proton (blue and yellow) in this artist's impression. Credit: JYU

A newly discovered pumpkin-shaped atomic nucleus that spits out a proton just after being formed could help scientists to understand how heavy elements are made in the Universe<sup>1</sup>.

### Access options

Subscribe to Nature+

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

\$29.99

monthly

[Subscribe](#)

Subscribe to Journal

Get full journal access for 1 year

\$199.00

only \$3.90 per issue

[Subscribe](#)

All prices are NET prices.

VAT will be added later in the checkout.

Tax calculation will be finalised during checkout.

Buy article

Get time limited or full article access on ReadCube.

\$32.00

[Buy](#)

All prices are NET prices.

### **Additional access options:**

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

*Nature* **604**, 10 (2022)

*doi:* <https://doi.org/10.1038/d41586-022-00889-8>

## **References**

1. Auranen, K. *et al.* *Phys. Rev. Lett.* **128**, 112501 (2022).

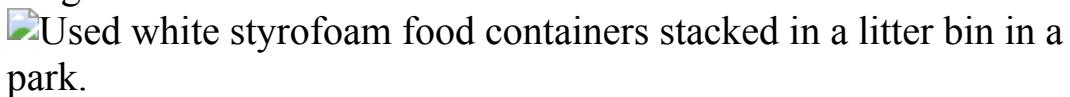
---

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-022-00889-8>

- RESEARCH HIGHLIGHT
- 31 March 2022

# A blast of light gives Styrofoam and other plastics new purpose

Polystyrene products can be recycled into a widely used compound thanks to a light-driven reaction.



Take-away boxes and other food packaging often contain the plastic polystyrene, which is difficult to reuse. Credit: Getty

A new processing method can transform waste polystyrene — a common type of plastic — into valuable small molecules<sup>1</sup>.

## Access options

Subscribe to Nature+

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

\$29.99

monthly

[Subscribe](#)

Subscribe to Journal

Get full journal access for 1 year

\$199.00

only \$3.90 per issue

## Subscribe

All prices are NET prices.

VAT will be added later in the checkout.

Tax calculation will be finalised during checkout.

Buy article

Get time limited or full article access on ReadCube.

\$32.00

## Buy

All prices are NET prices.

## **Additional access options:**

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

*Nature* **604**, 10 (2022)

*doi:* <https://doi.org/10.1038/d41586-022-00891-0>

## **References**

1. Oh, S. & Stache, E. E. *J. Am. Chem. Soc.* <https://doi.org/10.1021/jacs.2c01411> (2022).

---

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-022-00891-0>

- RESEARCH HIGHLIGHT
- 29 March 2022

# Secret of the birds: nearly 300 species fool foes the same way

The ‘broken wing’ trick has evolved in more than 50 bird families, hinting at its success in protecting eggs and chicks.

 A Semipalmated Plover (*Charadrius semipalmatus*) is doing a broken-wing display to distract predators from its nest.

The semipalmated plover (*Charadrius semipalmatus*) is one of many shorebirds that feigns injury to lure predators away from its nest. Credit: Chris Schenk/Buiten-beeld/Minden Pictures

A surprising number of bird species — at least 285 belonging to 52 families — playact being injured to protect their nests<sup>1</sup>.

## Access options

Subscribe to Nature+

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

\$29.99

monthly

[Subscribe](#)

Subscribe to Journal

Get full journal access for 1 year

\$199.00

only \$3.90 per issue

## Subscribe

All prices are NET prices.

VAT will be added later in the checkout.

Tax calculation will be finalised during checkout.

Buy article

Get time limited or full article access on ReadCube.

\$32.00

## Buy

All prices are NET prices.

## **Additional access options:**

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

*Nature* **604**, 11 (2022)

*doi:* <https://doi.org/10.1038/d41586-022-00890-1>

## **References**

1. De Framond, L., Brumm, H., Thompson, W. I., Drabing, S. M. & Francis, C. D. *Proc. R. Soc. B* <https://doi.org/10.1098/rspb.2022.0058> (2022).

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

- RESEARCH HIGHLIGHT
- 28 March 2022

# Marauding crazy ants come to grief when a fungus comes to call

Insect's populations defy many conventional pesticides but collapse after infection by a microbial pathogen.

A large group of tawny crazy ants swarm on a cobweb spider.

Tawny crazy ants, which can form dense infestations spanning kilometres in diameter, swarm a spider. Credit: Mark Sanders

Swarms of ‘crazy ants’ that invade houses, cause electrical short circuits and overrun birds’ nests might have met their match: a naturally occurring parasite<sup>1</sup>.

## Access options

Subscribe to Nature+

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

\$29.99

monthly

[Subscribe](#)

Subscribe to Journal

Get full journal access for 1 year

\$199.00

only \$3.90 per issue

[Subscribe](#)

All prices are NET prices.

VAT will be added later in the checkout.

Tax calculation will be finalised during checkout.

Buy article

Get time limited or full article access on ReadCube.

\$32.00

[Buy](#)

All prices are NET prices.

### **Additional access options:**

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

*Nature* **604**, 11 (2022)

doi: <https://doi.org/10.1038/d41586-022-00888-9>

## **References**

1. LeBrun, E. G., Jones, M., Plowes, R. M. & Gilbert, L. E. *Proc. Natl Acad. Sci. USA* **119**, e2114558119 (2022).

---

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-022-00888-9>

- RESEARCH HIGHLIGHT
- 31 March 2022

# Bacteria in China's livestock amass weapons against antibiotics

*Escherichia coli* sampled from pigs and other farm animals show a steep rise in antibiotic-resistance genes.

 Coloured scanning electron micrograph (SEM) of tube-shaped yellow bacteria attached to a rough-textured pink cell.

*Escherichia coli* bacteria (yellow; artificially coloured) swarm a dying cell.  
Credit: Steve Gschmeissner/Science Photo Library

Bacterial strains collected from livestock in China in the past few years had twice as many genes conferring resistance to antibiotics as did strains collected half a century earlier<sup>1</sup>.

## Access options

Subscribe to Nature+

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

\$29.99

monthly

[Subscribe](#)

Subscribe to Journal

Get full journal access for 1 year

\$199.00

only \$3.90 per issue

## Subscribe

All prices are NET prices.

VAT will be added later in the checkout.

Tax calculation will be finalised during checkout.

Buy article

Get time limited or full article access on ReadCube.

\$32.00

## Buy

All prices are NET prices.

## **Additional access options:**

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

*Nature* **604**, 11 (2022)

*doi:* <https://doi.org/10.1038/d41586-022-00892-z>

## **References**

1. Yang, L. *et al.* *Nature Food* <https://doi.org/10.1038/s43016-022-00470-6> (2022).

---

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-022-00892-z>

# News in Focus

- **Mystery space circle, coronavirus genomes and COVID antibody treatments** [ 06 April 2022]  
News Round-Up • The latest science news, in brief.
- **Exclusive: Documents reveal NASA's internal struggles over renaming Webb telescope** [ 25 March 2022]  
News • E-mails show agency's controversial response to astronomers concerned about past LGBT+ discrimination.
- **Will Omicron finally overpower China's COVID defences?**  
[ 28 March 2022]  
News • The country needs to control the virus until it has boosted vaccination rates in elderly people and reinforced the health-care system.
- **Biden bids again to boost science spending — but faces long odds** [ 28 March 2022]  
News • The US president wants huge increases for clean energy and public health, but a divided Congress might not go along with the plan.
- **Can drugs reduce the risk of long COVID? What scientists know so far** [ 25 March 2022]  
News • Researchers are trying to establish whether existing COVID-19 vaccines and treatments can prevent lasting symptoms.
- **Dozens of unidentified bat species probably live in Asia — and could host new viruses** [ 29 March 2022]  
News • Study suggests some 40% of horseshoe bats in the region have yet to be formally described.
- **Could computer models be the key to better COVID vaccines?** [ 05 April 2022]  
News Feature • For vaccine dosing decisions, past experience and best guesses won the day in the mad rush to beat back the pandemic. Modelling tools might have made a difference.
- **Why the WHO took two years to say COVID is airborne**  
[ 06 April 2022]  
News Feature • Early in the pandemic, the World Health Organization stated that SARS-CoV-2 was not transmitted through the air. That mistake and the prolonged process of correcting it

sowed confusion and raises questions about what will happen in the next pandemic.

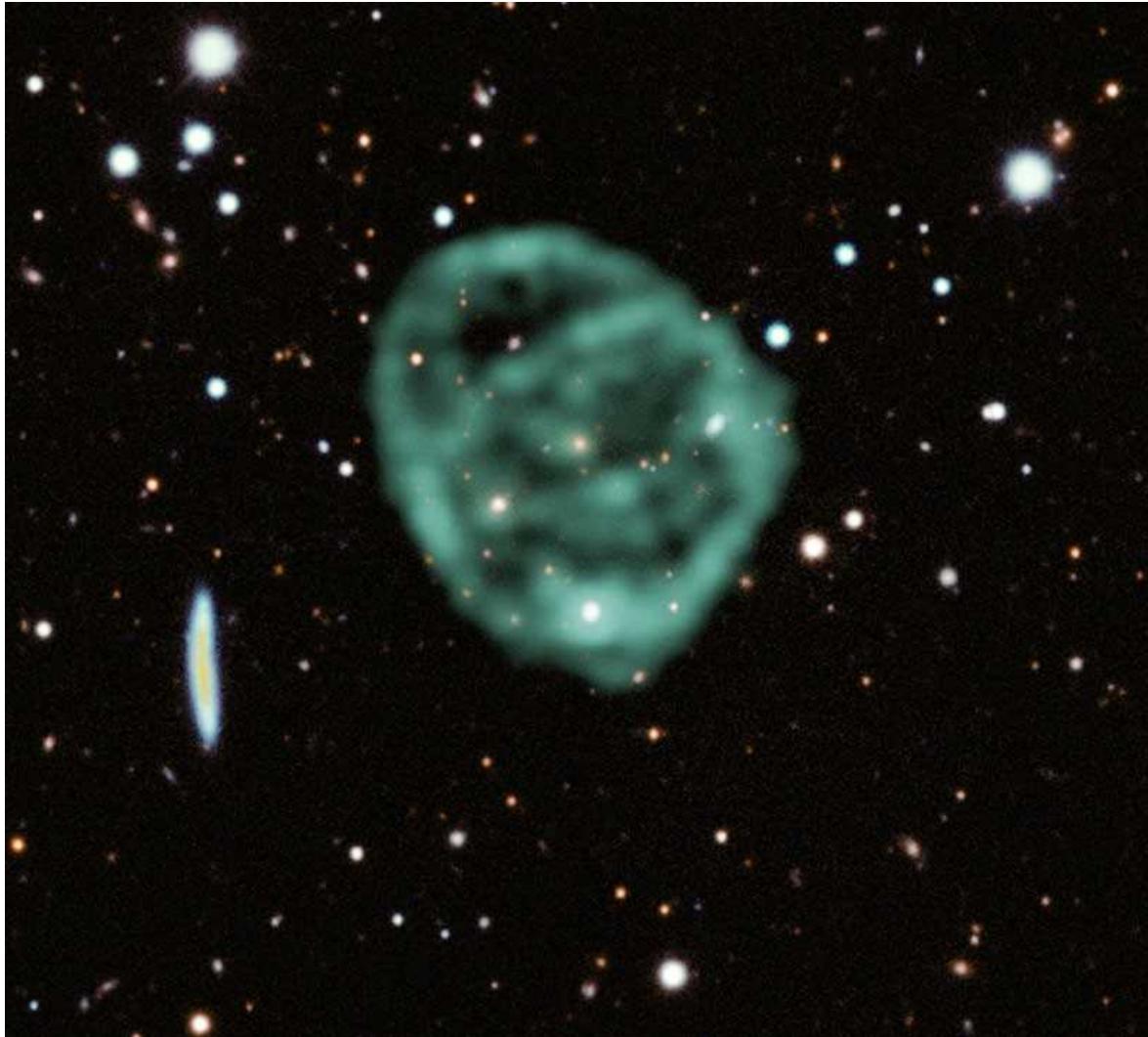
---

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

- NEWS ROUND-UP
- 06 April 2022

# Mystery space circle, coronavirus genomes and COVID antibody treatments

The latest science news, in brief.



The odd radio circle's large outer circle is possibly more than a million light years across. Credit: J. English (U. Manitoba)/EMU/MeerKAT/DES(CTIO)

## Bizarre space circle captured in unprecedented detail

Astronomers have captured a [close-up image of a rare and mysterious space object](#), prompting a renewed push to understand its origin. Odd radio circles (ORCs) are gigantic rings of radio waves. Only five have been sighted, and such spectacular detail is unequalled.

The image of ORC J2103-6200, also called ORC1, was captured by the high-resolution MeerKAT radio telescope in South Africa, which has given researchers unprecedented information about these rare phenomena ([R. P. Norris et al. Preprint at https://arxiv.org/abs/2203.10669; 2022](#)).

“This discovery will start new scientific research among astronomers,” says Alice Pasetto, a radio astronomer at the National Autonomous University of Mexico in Mexico City.

The MeerKAT radio data show that the ORC’s large outer circle could be more than 300,000 parsecs (one million light years) across — ten times the diameter of the Milky Way — with a series of smaller rings inside.

Researchers have proposed several theories to explain the origin of ORCs. These include that the objects are created by a shock wave from the centre of a galaxy, similar to what happens when two supermassive black holes merge. Another theory is that ORCs are shells caused by starbursts in the centre of a galaxy.

## Data gaps could slow hunt for next COVID variant

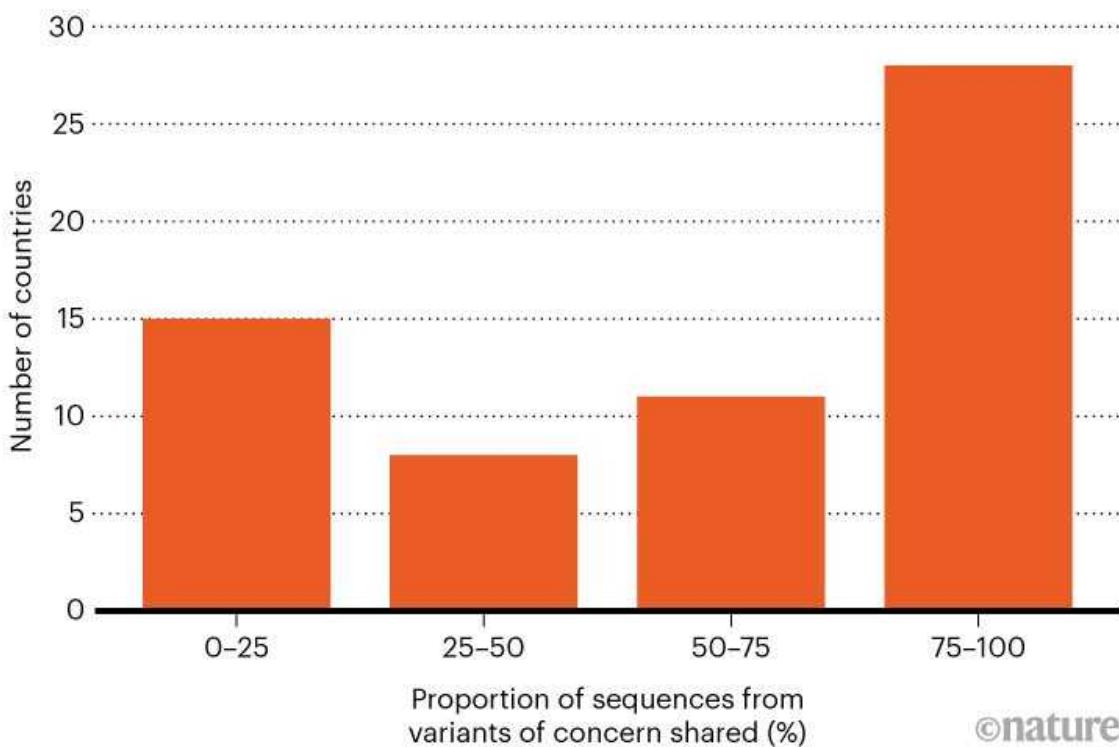
Many countries sequencing SARS-CoV-2 genomes are sharing only a fraction of them in public repositories — and many sequences lack

important information, according to a [global analysis of genomic surveillance](#).

Researchers are concerned that data gaps could make it harder to spot the next COVID-19 variant of concern, and could frustrate efforts to respond to it quickly.

## HIDDEN VARIANTS

Many countries shared fewer than half of their genome sequences from SARS-CoV-2 variants of concern.



Source: [Z. Chen et al. Nature Genet. https://doi.org/hn3x \(2022\)](#)

The authors of the analysis collected genomic data uploaded to public repositories including GISAID between the beginning of the pandemic and 31 October 2021 — comprising roughly 4.9 million genomes from 169 countries ([Z. Chen et al. Nature Genet. https://doi.org/hn3x; 2022](#)).

They compared the sequences with official reports from individual countries and found that of 62 countries that report such data, 23 — more than one-

third — had uploaded fewer than 50% of their sequences from the variants of concern Alpha, Beta, Gamma and Delta (see ‘Hidden variants’).



A tray holds supplies for administering monoclonal antibodies to treat COVID-19. Credit: Kayana Szymczak/Bloomberg/Getty

## Antibodies work better against COVID when given earlier

A first-of-its-kind analysis has a clear message about when to give the COVID-19 therapies called monoclonal antibodies: the earlier, the better ([E. Stadler et al. Preprint at medRxiv https://doi.org/hn9s; 2022](#)). The study has not yet been peer reviewed.

Monoclonal antibodies are synthetic versions of natural immune molecules that can prevent pathogens from infecting cells. They are among the few successful treatments for COVID-19.

To investigate the best timing and dose of these therapies, scientists at the University of New South Wales in Sydney, Australia, and elsewhere analysed data from an array of randomized controlled trials. They found that [early treatment could reduce the risk of hospitalization by around 70%](#). They

also determined that doses of antibody drugs 7 to 1,000 times lower than typically administered could still achieve nearly the same efficacy.

This finding could have far-reaching implications, given monoclonal antibodies' extraordinarily high cost and their ability to keep people with COVID-19 out of hospital.

*Nature* **604**, 13 (2022)

*doi:* <https://doi.org/10.1038/d41586-022-00923-9>

---

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-022-00923-9>

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

- NEWS
- 25 March 2022

# Exclusive: Documents reveal NASA's internal struggles over renaming Webb telescope

E-mails show agency's controversial response to astronomers concerned about past LGBT+ discrimination.

- [Alexandra Witze](#)



Shown here being packed up for shipment to its launch site last year, the James Webb Space Telescope was designed to probe the early Universe.Credit: Chris Gunn/NASA

Internal NASA documents obtained by *Nature* reveal fresh details about [the agency's investigation last year](#) into whether to rename its flagship James Webb Space Telescope (JWST). A group of astronomers had led a community petition to change the name, alleging that the telescope's namesake, former NASA chief James Webb, had been complicit in the persecution and firing of gay and lesbian federal employees during his career in the US government in the 1950s and 1960s.

In September, [NASA announced that it would not change the name](#) of the telescope, because it had no evidence to support the allegations. But the agency controversially did not release a report summarizing its investigation or decision-making.

The internal documents obtained by *Nature* and others through freedom-of-information (FOI) requests show that, while making its decision, the agency was aware of a 1969 appeals ruling suggesting that it had been customary at NASA to fire people over suspicions about their sexual orientation. The case involved a former NASA employee who had been fired in 1963 because supervisors thought he was gay. This was when Webb was leading the agency.

NASA's acting chief historian, Brian Odom, says he has not found any information in NASA's archives to suggest that firing people for their sexual orientation was agency policy under Webb. He and a contract historian, whose identity has not been disclosed, will soon visit other historical archives to continue to look into Webb's history. These archives have been closed owing to the COVID-19 pandemic but will open again in the next few weeks. Odom says he will share information about what the historians find with the astronomy community.

But the new FOI documents shed light on how NASA has looked into the matter so far. The e-mail correspondences "paint a stark portrait of how astronomers outside the LGBTQ+ community dismiss the experiences of their queer colleagues, and make it plain to see that discrimination against

queer people is alive and well in astronomy today”, say the four astronomers who led the community petition. They are Lucianne Walkowicz at the Adler Planetarium in Chicago, Illinois; Chanda Prescod-Weinstein at the University of New Hampshire in Durham; Brian Nord at the Fermi National Accelerator Laboratory in Batavia, Illinois; and Sarah Tuttle at the University of Washington in Seattle.

## Agency influencer

Webb ran NASA between 1961 and 1968, when the Apollo programme to send astronauts to the Moon was at the height of its development. He died in 1992. Another former NASA administrator, Sean O’Keefe, named the telescope after Webb in 2002 to recognize his leadership in government and his commitment to making science a key part of the agency. “We wouldn’t have gotten to the Moon when we did in 1969 without his leadership,” says Barbara Webb, who is James Webb’s daughter-in-law. “And the telescope is an amazing instrument, the greatest scientific and engineering instrument ever created — it’s very fitting that it’s named for him.”

Barbara Webb says her family does not believe James Webb discriminated against anyone. “I don’t think that saying he was a bigot is accurate in any sense,” she says. “He had more integrity than anybody I ever knew.”

Some astronomers have argued that whatever Webb’s personal beliefs, the telescope should be renamed because he was a top-ranking official — and therefore had influence — at a time when the US federal government systematically identified and fired employees because of suspicions about homosexuality. This period of history, known as the lavender scare, began at the US Department of State around 1950, when Webb was the number-two official there.



James Webb was head of NASA from 1961 to 1968.Credit: Bettmann/Getty

The controversy ties into [broader conversations](#) about the implications of naming buildings, spacecraft and other objects after people or concepts with complicated histories. [Launched in December](#), the US\$10-billion JWST is a landmark observatory supported by NASA, as well as the European and Canadian space agencies. The telescope, which is expected to make its first scientific observations no earlier than June, is designed to [peer at galaxies near the dawn of time](#) and explore other cosmic frontiers. Its name will dominate astronomical publications for years to come.

Current NASA administrator Bill Nelson made the choice to keep the name. He has not provided any further statements.

## Troubling findings

The new FOI documents are only a partial glimpse into NASA's decision-making (see the PDF in Supplementary Information below; *Nature* additionally redacted details on a former NASA intern to preserve the

person's privacy). They encompass some of the correspondence between NASA employees about the agency's investigation into Webb from 1 January to 13 October of last year. *Nature* filed its request because the agency had released few details about its September decision.

Although the documents reveal that key decisions were made in meetings and not over e-mail, they still show agency officials wrestling with how to investigate the allegations and control public messaging over the controversy. As early as April 2021, an external researcher flagged wording from the 1969 court ruling to NASA officials. It came in the case of Clifford Norton, who had appealed against being fired from NASA for "immoral, indecent, and disgraceful conduct". In the decision, the chief judge wrote that the person who had fired Norton had said that he was a good employee and asked whether there was a way to keep him on. Whomever he consulted in the personnel office told him that it was a "custom within the agency" to fire people for "homosexual conduct".

"I think you will find this paragraph to be troubling," wrote the external researcher to Eric Smith, the JWST's programme scientist at NASA in Washington DC. "'A custom within the agency' sounds pretty bad."

A white paper drawn up within NASA, and described as not meant for public release, says: "This shows that NASA had decided that removal of homosexual employees would be its policy. They had a choice during Webb's tenure as administrator to set or change that policy."

Also in April, the e-mails show, Paul Hertz, the head of NASA's astrophysics division, contacted more than ten members of the astrophysics community to ask their opinions about possibly changing the telescope's name. "Nobody said they would be disappointed if we did not change the name," Hertz wrote to his manager. However, none of them were known members of the LGBT+ community, Hertz also wrote.

When asked about this by *Nature*, Hertz said he has had numerous conversations on the issue with "members of many communities, including those from LGBTQ+".

## A lack of rationale

In October, after NASA announced that it would not rename the telescope, Hertz summarized community reaction in an e-mail to another NASA employee: “The problem for most of the astrophysics community is not the decision itself, but the lack of rationale to explain why this is the appropriate decision. (For some people, the problem is the decision.)” He also pointed out that NASA had not transmitted the decision broadly, but rather e-mailed Nelson’s one-sentence statement to a small group of news outlets, including *Nature*. “That is hardly the kind of transparent process that gives the external community confidence,” he wrote.

Walkowicz and their colleagues go much further. “This is harm from a community we have committed substantive time to and which we are passionate about,” they say.

In November, the Astrophysics Advisory Committee, a group of independent researchers who advise NASA, [asked Hertz for a written report on NASA's decision](#). “The rationale and explanation from the Agency were wholly insufficient, non-transparent, and uninformative regarding the naming decision communicated to the astronomical community and other stakeholders regarding the James Webb Space Telescope,” the committee wrote. It has not yet received such a report.

NASA states on its website and elsewhere that it supports diversity, equity, inclusion and accessibility. “NASA is entirely committed to the full participation and empowerment of a wide variety of people, organizations, capabilities, and assets because we know this best enables us to access everyone and everything we need to best accomplish our missions,” its [policy](#) states.

The revelations about NASA’s decision regarding the JWST come at a time of increasing concern over the way the agency handles issues of identity. Earlier this month, employees at the agency’s Goddard Space Flight Center in Greenbelt, Maryland, [were told that they would no longer be able to include pronouns](#), such as she/her or they/them, in their display names in agency computer systems. After the move was discussed on Reddit and the astronomy community reacted negatively on other social platforms, [NASA put out a statement that employees could continue](#) to include pronouns in their e-mail signature blocks.

*Nature* **604**, 15-16 (2022)

*doi:* <https://doi.org/10.1038/d41586-022-00845-6>

---

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-022-00845-6>

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

- NEWS
- 28 March 2022
- Correction [30 March 2022](#)

# Will Omicron finally overpower China's COVID defences?

The country needs to control the virus until it has boosted vaccination rates in elderly people and reinforced the health-care system.

- [Dyani Lewis](#)



In some parts of China, barricades have been set up around communities with high numbers of COVID-19 cases. Credit: Kevin Frayer/Getty

All eyes are on China as it attempts to quash its largest COVID-19 outbreaks since the early days of the pandemic. More than 62,000 people across all 31 of its provinces are infected, most of them with the fast-spreading BA.2 Omicron variant.

The outbreaks have plunged tens of millions of people into lockdown. President Xi Jinping announced earlier this month that China would stick to its ‘dynamic zero-COVID strategy’, which aims to stamp out infections and prevent the virus from spreading through communities. This policy now stands in contrast to a global trend towards easing restrictions and attempting to co-exist with the circulating virus.

China’s hard-line approach to eliminating COVID-19 seems to be softening. In his speech, Xi also flagged a more pragmatic strategy, asking that officials limit the economic impact of control measures. In practice, this means that people with asymptomatic cases of COVID-19 are being sent to dedicated isolation centres rather than hospitals, and are monitored for shorter periods than previously required. But some researchers are divided about whether the virus will spread out of control before the government has time to prepare.

## Testing boost

“They’ve demonstrated time and time again that they can control outbreaks,” says epidemiologist Ben Cowling at the University of Hong Kong. The country successfully controlled the hyper-transmissible Omicron variant during the Beijing winter Olympics in February, despite predictions that large outbreaks would occur, he says. He anticipates that the aggressive testing efforts under way will see case numbers continue to rise over the next few days, but that numbers will fall back to zero after that.

If cases do drop, the government will probably spend the next year boosting the low vaccination rates in elderly people and establishing primary health-care infrastructure to ease pressure on hospitals, says economist Xi Chen at Yale University in New Haven, Connecticut, who studies China’s public-health system.

But others say that China needs to brace itself for a growing outbreak. New infections are edging towards 6,000 confirmed cases a day. “It’s growing quickly — I think it’s out of control now,” says Michael Osterholm, an infectious-diseases epidemiologist at the University of Minnesota in Minneapolis. Case numbers are unlikely to come back down to zero, he says, and trying to keep the virus suppressed would require such severe restrictions that “it surely will bring down their economy”.

If Omicron runs out of control, the effects could be devastating — and similar to the current outbreak in Hong Kong, where deaths have surged and hospitals are overwhelmed. An analysis by Airfinity, a life-sciences market analytics firm in London, suggests that more than one million people in mainland China could die during an Omicron wave, partly because of lower levels of protection in older people. Only 50% of people over the age of 80 are fully vaccinated, according to the Chinese government.

## Lessons from Hong Kong

Recent experience in Hong Kong has highlighted the cost of low vaccination rates in older people. In early March, there were close to 900 cases of COVID-19 per 100,000 residents in Hong Kong, the highest level recorded anywhere in the world during the pandemic. Deaths have also surged to nearly 300 a day earlier this month. Experts blame low vaccination rates in elderly people for the region’s high mortality rate. Only about one-third of those aged over 80 years are fully vaccinated, and 90% of deaths have been in people who are not fully vaccinated.

Mainland China faces a similar predicament if the current outbreak is not controlled. China’s overall vaccination rate is higher than 85%. That has been achieved with the introduction of a digital vaccination-passport system — required for entry into many public buildings and workplaces — and a colour-coded ‘health code’ that indicates whether someone poses an infection risk. But Cowling says that older people are less likely to use facilities that require a vaccination passport and have been able to remain unvaccinated.

Fifty-two million people aged over 60 years are yet to be fully vaccinated. The most vulnerable — those aged over 80 years — are the least well vaccinated, with only 20% having received the primary vaccination course and booster shot. Work by Cowling and his colleagues that is yet to be peer reviewed indicates that the Sinovac vaccine, one of the two main vaccines used in China, is effective at reducing severe cases and deaths, but that the third shot is necessary to confer high levels of protection in those aged over 60<sup>1</sup>.

Deaths from COVID-19 could be “much worse” than the 1 million estimated by Airfinity if Omicron spreads through the population, says Lu Jiahai, an infectious-diseases epidemiologist at Sun Yat-sen University in Guangzhou, China. The government is “taking responsibility for people’s lives, so will not change or relax the current prevention and control strategy”, he says.

## Exit strategy

China’s large domestic economy has made a zero-COVID strategy more sustainable than it has been in places with smaller populations, such as New Zealand and Singapore. But it can’t last forever. The stock market index is falling, suggesting that China’s economy is suffering, and Cowling says that disruptive outbreaks will become more frequent as the virus circulates freely elsewhere.

Boosting vaccination rates in elderly people will be a priority, but many live in rural areas, so this will take time. Lu says that China will probably wait until the vaccination rate reaches 80% in that group before easing restrictions further.

Chen says that, at the moment, there are few family doctors in China, which means people rely on hospitals as their first port of call. He suspects that most COVID-19 restrictions will be in place for another year. This would allow the government to build up a fledgling family doctor network or dedicated isolation facilities to treat people with mild or asymptomatic COVID-19 and ease pressure on the hospital system. If the country isn’t prepared, re-opening could be “a disaster in terms of the health-care system”, says Chen.

Chinese authorities also need to prepare the public for what's to come, says Chen. Many people have come to rely on the government's willingness to use all means necessary to control the virus at the expense of the economy, he says.

*Nature* **604**, 17-18 (2022)

*doi:* <https://doi.org/10.1038/d41586-022-00884-z>

## Updates & Corrections

- **Correction 30 March 2022:** An earlier version of this story incorrectly attributed the rates of vaccination in people over 80 on the Chinese mainland to an Airfinity report rather than the Chinese government. This has been corrected.

## References

1. McMenamin, E. *et al.* Preprint at medRxiv  
<https://doi.org/10.1101/2022.03.22.22272769> (2022).

---

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-022-00884-z>

- NEWS
- 28 March 2022
- Correction [30 March 2022](#)

# Biden bids again to boost science spending — but faces long odds

The US president wants huge increases for clean energy and public health, but a divided Congress might not go along with the plan.

- [Jeff Tollefson](#) ,
- [Max Kozlov](#) ,
- [Amy Maxmen](#) &
- [Alexandra Witze](#)



“We are tackling the climate crisis with urgency”, and “strengthening the global health architecture to combat COVID-19 and future pandemics”, US President Joe Biden wrote in his budget proposal for 2023, released on 28 March. Credit: Slawomir Kaminski/Agencja Wyborcza via Reuters

US President Joe Biden [released his latest budget request](#) to Congress on 28 March, calling for significant new investments in clean energy and public health. The proposal for fiscal year 2023 includes roughly US\$200 billion for research and development programmes, according to the American Association for the Advancement of Science in Washington DC. It comes just weeks after Congress finalized a spending package for the current fiscal year that cut back considerably on the science [funds that Biden had requested](#) — signalling more challenges ahead.

Although Congress modestly increased funding for most science agencies in 2022, it heavily scaled back some of the administration’s most ambitious proposals. For instance, the Advanced Research Projects Agency—Health (ARPA-H), a new high-risk, high-reward agency meant to accelerate innovations in health and medicine, received \$1 billion rather than the \$6.5 billion the president requested last year.

“The president’s request is certainly very solid, and the increases for science go right across the board,” says Michael Lubell, a physicist at the City College of New York who tracks federal science-policy issues. But Lubell says concerns about the rising cost of living owing to inflation, as well as economic uncertainty around Russia’s invasion of Ukraine, could make it hard for Biden’s party, the Democrats, to boost government spending on anything, without also putting new taxes in place to increase revenue. “Science is one of the least contentious issues,” Lubell adds, “but it’s not something that puts food on the table tomorrow.”

## Health priorities

A standout in Biden’s budget proposal is the request for a mandatory \$81.7 billion over five years to prepare the United States for future pandemics and other biological threats. “This new pandemic-preparedness effort is designed to address the deficiencies with our public-health system that we have

witnessed” in the past few years, says Jennifer Kates, director of global health and HIV policy at KFF, a non-profit health-policy organization headquartered in San Francisco, California. Kates, however, notes that the pitch will probably struggle to get approval in Congress, both because it’s a large financial demand and because the funds would be mandatory until 2027, as opposed to being at the discretion of policymakers each year. In making the budget mandatory, the administration is hoping to avoid the cycle of panic and neglect that often follows health emergencies, Kates says.

More than one-third of the preparedness funds are slated for the US Centers for Disease Control and Prevention (CDC). On top of the normal budget increase proposed for the agency, this would mean a whopping 171% boost in funds for the CDC compared with 2021. (The White House has not yet released its official budget numbers for 2022.) The budget request for the agency includes funds earmarked for coronavirus sequencing and surveillance, for efforts to modernize public-health data collection and for the CDC’s new Center for Forecasting and Outbreak Analytics, where disease-modelling specialists will analyse data that can help to guide public-health decisions.

Under Biden’s plan, the National Institutes of Health (NIH) would also receive a dramatic increase in funds, with much of the jump related to pandemic preparedness. The biomedical agency would get a total of \$62.5 billion, representing a 46% increase from 2021. Of that, \$12.1 billion would go towards researching and developing vaccines, diagnostics and therapeutics against pathogens of pandemic concern, and to expanding laboratory and clinical-trial infrastructure.

Another \$5 billion of the NIH’s proposed budget would go to ARPA-H in 2023. Although ARPA-H is included in the NIH’s budget request, policymakers continue to spar over whether the health-innovations agency should be housed in the NIH, or whether it should be a separate branch of the Department of Health and Human Services. [Some think that the NIH is too bureaucratic and conservative](#) in its approach to biomedical research.

However, without the sums budgeted for pandemic preparedness and ARPA-H, the NIH would actually receive only \$275 million more than Congress allocated for the agency in 2022, says Jennifer Zeitzer, the director of public

affairs at the Federation of American Societies for Experimental Biology in Rockville, Maryland. Because the NIH is the largest public funder of biomedical research in the world, Zeitzer says, this level of funding is inadequate. “It would be very unusual for a Democratic president — or any president — to not propose [a more substantive] baseline increase for NIH,” she adds.

Even so, the Biden administration proposed modest increases to several research areas, including those related to mental and maternal health, health disparities and inequities, and sexual orientation and gender identity. The request also proposes more funding for the [Cancer Moonshot initiative](#), a programme that is attempting to reduce the cancer death rate by at least 50% over the next 25 years.

The request would also boost the Food and Drug Administration’s (FDA’s) budget by nearly 39% compared with 2021 — to \$8.4 billion — largely owing to a \$1.6-billion boost from the pandemic-preparedness fund that would allow the agency to modernize regulatory capacity, speed development of diagnostics and strengthen the supply chain for personal protective equipment.

## Energy and climate spending

Biden has made climate change [a top priority of his administration](#), and proposed historic government investments in clean-energy innovation, but Republicans and some Democrats have stymied much of that agenda on Capitol Hill. In this year’s budget proposal, the president is once again seeking significant increases for energy, environment and climate science across the federal government.

The Environmental Protection Agency (EPA) would receive \$11.9 billion, an increase of 29% compared with 2021. The National Oceanic and Atmospheric Administration would see its budget increase by more than 25% compared with 2021, to \$6.9 billion, which includes \$2.3 billion for weather satellites.

But the biggest energy and climate investments are at the Department of Energy (DoE), which runs a suite of applied-energy research and technology demonstration programmes. The department received a one-time boost of \$21.5 billion as part of a larger infrastructure package enacted by Congress last November; that money is dedicated to commercial demonstration programmes focused on a range of energy sources, from hydrogen to nuclear power. Coming on top of that, the latest budget proposal would provide \$48.2 billion for the agency, a 15% increase compared with 2021 levels, with broad increases for research and development into renewable energy sources such as solar and wind power, as well as carbon capture and other low-carbon technologies.

All told, the president's budget would provide a record \$19.1 billion for the DoE's main clean-energy programmes over the coming year, says David Hart, who heads the Center for Clean Energy Innovation at the Information Technology and Innovation Foundation, a think tank based in Washington DC. Around half of that money is already locked in through the infrastructure law, which also allows the DoE to spread its spending over five years.

In the end, Hart says, the DoE might even struggle to spend all of the money it has received. "It's going to be a challenge for them to spend it wisely," he says.

In parallel, the administration is still pushing for an estimated \$550 billion in funding for climate and energy initiatives that were included in a [\\$2-trillion piece of legislation](#) called the Build Back Better bill. That hit a wall in the US Senate, however, because of united opposition from Republicans and broken ranks among Biden's fellow Democrats. The question now is whether Democrats and the administration will be able to revive the climate provisions, which had broad support in the party.

## Other science investments

Biden's proposal also includes a request to increase the budget of the National Science Foundation (NSF) by 23.6% compared with 2021, to a total of around \$10.5 billion. The NSF funds a significant chunk of US

academic research, and is considered by many researchers to be underfinanced. Last year, Biden requested a 20% increase to the NSF budget, but it was cut down by Congress to about a 4.1% increase.

Biden's latest request falls in line with the administration's broader priorities. Of the additional \$2 billion the agency would receive, around \$1.6 billion is allocated to climate-resilience research. The request also sets aside \$880 million for [the NSF's technology directorate](#) — the agency's first new directorate in around 30 years. The Technology, Innovations and Partnerships Directorate was established in early March in an effort to sharpen the United States' competitive edge globally, in economically important technologies such as superconductors and artificial intelligence.

The Senate and House of Representatives are [continuing to nail down](#) the details of the directorate with duelling bills, slated to be reconciled sometime in the coming months. Still, this is an indication that Congress has renewed interest in scientific research, says Zeitzer, adding that a 4.1% increase is “not terrible” for the NSF’s budget. But others worry that Biden’s 2023 request is still too low.

“In most years, a 20-plus per cent increase would be considered a very good budget” request, says David Cheney, a managing partner at Technology Policy International in Los Osos, California. “But it’s a little bit disappointing to people who have looked at the House and Senate authorization bills, which envisioned larger increases.”

For NASA, the White House requests \$26 billion, which would be an 11.6% increase from 2021. Much of that is focused on the agency’s Artemis initiative to return humans to the surface of the Moon no earlier than 2025. Science budgets at NASA would rise by more than 9% over 2021 levels, but much of that increase would be concentrated in the Earth-science division. The agency’s astrophysics division would see a 12% drop, even as it gears up to work on ambitious space-based missions laid out in [a community road map](#) late last year.

One notable space-related item, in the Department of Commerce request, is that Biden is asking Congress for \$88 million, a \$78 million increase over 2021, for the fledgling Office of Space Commerce, which is intended to

track the locations of active satellites and space debris. [Thousands of satellites have been launched](#) in the past few years, and space analysts have been calling for the United States to respond more effectively to the growing congestion in low Earth orbit.

Congress will now consider Biden's requests and hammer out the details of how much to allocate to each agency.

*Nature* **604**, 18-19 (2022)

*doi:* <https://doi.org/10.1038/d41586-022-00897-8>

## Updates & Corrections

- **Correction 30 March 2022:** An earlier version of this story said that Congress cut Biden's requested boost for the NSF's fiscal 2022 budget from about 20% to 4.3%. It cut the boost to 4.1%.

---

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-022-00897-8>

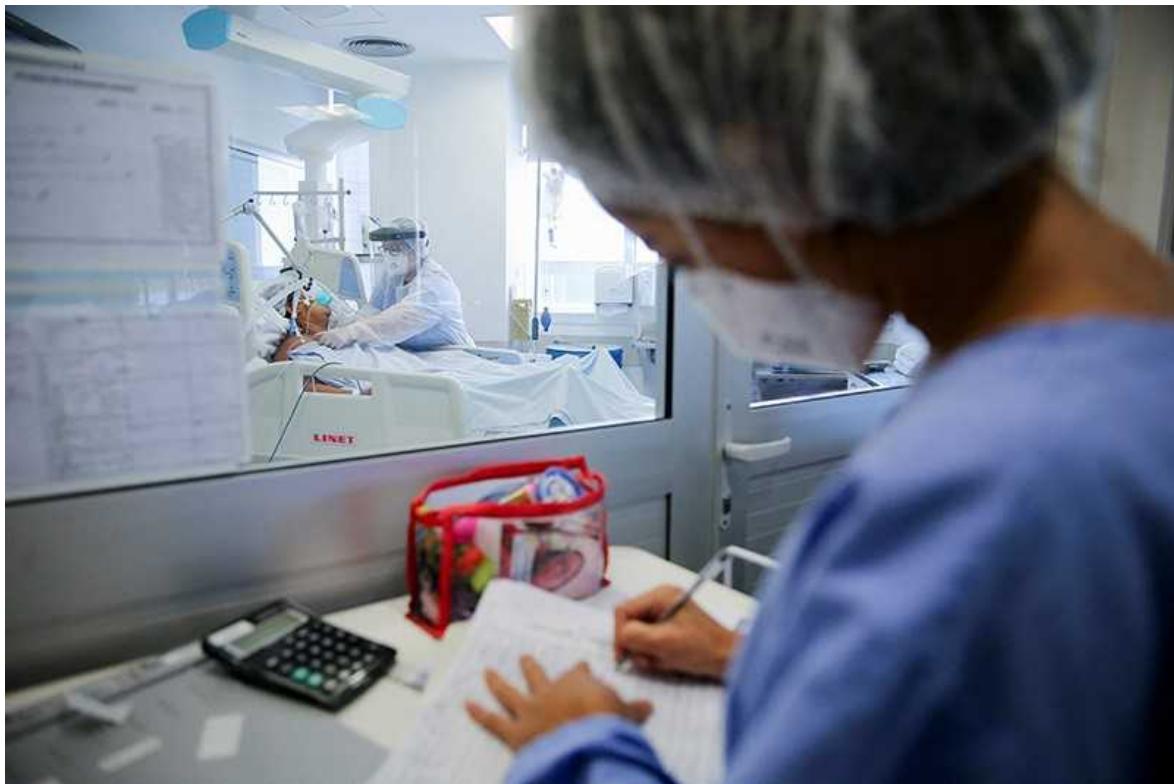
| [Section menu](#) | [Main menu](#) |

- NEWS
- 25 March 2022

# Can drugs reduce the risk of long COVID? What scientists know so far

Researchers are trying to establish whether existing COVID-19 vaccines and treatments can prevent lasting symptoms.

- [Heidi Ledford](#)



Researchers are studying the long-term effects of treatments given to people hospitalized with COVID-19. Credit: Diego Vara/Reuters

In the early days of the COVID-19 pandemic, urologist and clinical epidemiologist Kari Tikkinen found his schedule full of cancelled surgeries, so he had some time to kill. “Do whatever you feel is most useful,” his boss at the University of Helsinki advised him. So Tikkinen threw himself into running clinical trials for COVID-19 therapies.

From the start — before the world learnt of long COVID — Tikkinen saw a need to follow study participants for months after their recovery. He wanted to monitor long-term side effects of the medicines. “Very soon, it became clear: it’s not only about safety,” he says.

Now, Tikkinen and a handful of others are hoping to learn more about whether treatments given during the acute phase of COVID-19 can reduce the risk of experiencing symptoms months later. “It’s an urgent and pressing health need that people need to start focusing on,” says intensive-care specialist Charlotte Summers, at the University of Cambridge, UK.

## Debilitating symptoms

Research into long COVID — which is also known as post-acute sequelae of COVID-19, and is usually defined as COVID-19 symptoms that last longer than three months — has lagged behind studies of the acute phase of infection. People who experience long COVID live with a wide array of symptoms, ranging from mild to severely debilitating. Researchers have proposed a variety of causes for the condition — from lingering viral reservoirs, to autoimmunity, to tiny blood clots. Many think that a mix of these factors is to blame. “It took a while to get going on any serious mechanistic long-COVID research,” says immunologist Danny Altmann at Imperial College London. “It’s hard to piece the big picture together.”

Thus far, vaccines are the best way to prevent long COVID. COVID-19 vaccines reduce the risk of SARS-CoV-2 infection, and they might lessen the risk of long COVID after a breakthrough infection in someone who has been vaccinated.

Several studies have looked at this question: although they have yielded divergent results, the overall trend suggests that vaccination could reduce the

risk of long COVID by about half among those who become infected after vaccination. For example, one study<sup>1</sup> that has not yet been peer reviewed found that vaccination reduced the chances of developing long-COVID symptoms by about 41% in more than 3,000 double-vaccinated participants who were later infected with SARS-CoV-2.

But that still leaves too many people at risk of getting long COVID, says Altmann. “Half is not as good as I thought it would be,” he says. “I was thinking and hoping that long COVID would be a thing of the past.”

## Early treatment

Beyond vaccination, it’s unclear whether any existing COVID-19 therapy has an effect on long-COVID risk. In theory, a drug that reduces disease severity might reduce the severity of long-term symptoms, says Altmann. But long COVID is not always associated with serious acute illness. “There are loads of people out there who are really destroyed by long COVID and had asymptomatic or near asymptomatic infections,” he says. “It’s really hard to grapple with.”

Nevertheless, some studies plan to look at the impact of early treatment with antiviral drugs on long COVID. A clinical trial called PANORAMIC has been testing the effects of the oral antiviral molnupiravir, developed by Merck in Kenilworth, New Jersey, and Ridgeback Biotherapeutics in Miami, Florida, on COVID-19 severity. Although it is not the primary goal of the study, researchers will gather data from participants at three and six months after treatment — which could determine whether the drug affects long-COVID risk. Similarly, two trials of Paxlovid, an antiviral drug developed by Pfizer in New York City, will include a six-month follow-up of participants.

These antiviral drugs are typically used to treat people with relatively mild COVID symptoms. Tikkinen and his colleagues hope to learn more about the long-term impact of treatments received by those who were hospitalized with COVID-19. His team is following up with participants in the University of Helsinki’s arm of the World Health Organization’s international [COVID-19 treatment trial, called SOLIDARITY](#). In the next few weeks, he hopes to

have the results of a one-year follow-up study of participants who were hospitalized with COVID-19 and treated with the antiviral drug remdesivir.

His team will also follow up with participants in two other arms of the SOLIDARITY trial, one that tested an immune-suppressing drug called infliximab and another that tested imatinib, a drug that could help to reduce inflammation in blood vessels.

But, Tikkinen cautions, none of these studies had enough participants to give clear-cut answers on long COVID. His team went to extraordinary measures to contact participants months after their remdesivir treatment and to encourage them to fill out a survey about their symptoms. The team hired graphic designers to make the surveys easier to fill out, had the questions translated into ten languages and offered to hand-deliver the paperwork to participants' homes. The result was a 95% response rate, which Tikkinen says is unusually high for such long-term studies. But because the original study included only about 350 people, it is probably still too small to provide a definitive conclusion.

## **Small-scale trials**

Researchers are hoping to find out whether more treatments can reduce the risk of long COVID. A large UK-based trial called HEAL-COVID is testing two drugs that target the cardiovascular system in people who have been hospitalized with COVID-19. One, called apixaban, is an anticoagulant. The other, atorvastatin, is a cholesterol-lowering medication thought to reduce inflammation in blood vessels.

The study will investigate whether either treatment reduces hospitalizations and deaths in the year after people are first discharged from hospital. Nearly one-third of people who are discharged after treatment for COVID-19 are readmitted within six months, and 12% die within six months of their initial discharge. "And when we looked at what was most plausibly leading to death after hospitalization, it was probably the cardiopulmonary effects," says Summers, who is leading the study.

At the University of Chicago in Illinois, pulmonologist and critical-care physician Ayodeji Adegunsoye has observed a possible increase in the accumulation of scar tissue, called fibrosis, in the lungs well after the acute infection in people who were hospitalized with COVID-19 and required supplemental oxygen. He is now testing a drug called sirolimus — an immune-suppressing drug that is sometimes given to organ-transplant recipients — in such people, in the hope that it will prevent the migration of cells that promote fibrosis in the lung.

By their nature, long-COVID studies require patience: one commonly accepted definition of long COVID is the persistence of certain symptoms for more than 12 weeks after the acute infection. Altmann is optimistic that this year will yield advances, but cautions against reading too much into small trials that might not yield statistically meaningful results. “There’s such pressure,” he says. “There’s this incredibly pressing and desperate need — we all feel that anxiety.”

*Nature* **604**, 20-21 (2022)

*doi:* <https://doi.org/10.1038/d41586-022-00823-y>

## References

1. Ayoubkhani, D. *et al.* Preprint at medRxiv  
<https://doi.org/10.1101/2022.02.23.22271388> (2022).

---

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-022-00823-y>.

- NEWS
- 29 March 2022

# Dozens of unidentified bat species probably live in Asia — and could host new viruses

Study suggests some 40% of horseshoe bats in the region have yet to be formally described.

- [Smriti Mallapaty](#)



There could be more species of horseshoe bat than previously thought. Credit: Chien Lee/Nature Picture Library

A genomic analysis suggests that there are probably dozens of unknown species of horseshoe bats in southeast Asia<sup>1</sup>.

Horseshoe bats (Rhinolophidae) are considered the reservoir of many zoonotic viruses — which jump from animals to people — including the close relatives of the viruses that caused severe acute respiratory syndrome and COVID-19. Identifying bat species correctly might help pinpoint geographical hotspots with a high risk of zoonotic disease, says Shi Zhengli, a virologist at the Wuhan Institute of Virology in China. “This work is important,” she says.

The study was published in *Frontiers in Ecology and Evolution* on 29 March.

Better identification of unknown bat species could also support the search for the origins of SARS-CoV-2 by narrowing down where to look for bats that may harbour close relatives of the virus, says study co-author Alice Hughes, a conservation biologist at the University of Hong Kong. The closest known relatives of SARS-CoV-2 have been found in *Rhinolophus affinis* [bats in Yunnan](#) province, in southwestern China<sup>2</sup>, and in three species of [horseshoe bat in Laos](#)<sup>3</sup>.

## Cryptic species

Hughes wanted to better understand the diversity of bats in southeast Asia and find standardized ways of identifying them. So she and her colleagues captured bats in southern China and southeast Asia between 2015 and 2020. They took measurements and photographs of the bats’ wings and noseleaf — “the funky set of tissue around their nose”, as Hughes describes it — and recorded their echolocation calls. They also collected a tiny bit of tissue from the bats’ wings to extract genetic data.

To map the bats’ genetic diversity, the team used mitochondrial DNA sequences from 205 of their captured animals, and another 655 sequences from online databases — representing a total of 11 species of Rhinolophidae. As a general rule, the greater the difference between two

bats' genomes, the more likely the animals represent genetically distinct groups, and therefore different species.

The researchers found that each of the 11 species was probably actually multiple species, possibly including dozens of hidden species across the whole sample. Hidden, or 'cryptic', species are animals that seem to belong to the same species but are actually genetically distinct. For example, the genetic diversity of *Rhinolophus sinicus* suggests that the group could be six separate species. Overall, they estimated that some 40% of the species in Asia have not been formally described.

"It's a sobering number, but not terribly surprising," says Nancy Simmons, a curator at the American Museum of Natural History in New York City. Rhinolophid bats are a complex group and there has been only a limited sampling of the animals, she says.

However, relying on mitochondrial DNA could mean that the number of hidden species is an overestimate. That is because mitochondrial DNA is inherited only from the mother, so could be missing important genetic information, says Simmons. Still, the study could lead to a burst of research into naming new bat species in the region, she says.

## Further evidence

The findings corroborate other genetic research suggesting that there are many cryptic species in southeast Asia, says Charles Francis, a biologist at the Canadian Wildlife Service, Environment and Climate Change Canada, in Ottawa, who studies bats in the region. But, he says, the estimates are based on a small number of samples.

Hughes' team used the morphological and acoustic data to do a more detailed analysis of 190 bats found in southern China and Vietnam and found that it supported their finding that many species had not been identified in those regions. The study makes a strong argument for "the use of multiple lines of evidence when delineating species", says Simmons.

Hughes says her team also found that the flap of tissue just above the bats' nostrils, called the sella, could be used to identify species without the need

for genetic data. Gábor Csorba, a taxonomist at the Hungarian Natural History Museum in Budapest, says this means that hidden species could be identified without doing intrusive morphology studies or expensive DNA analyses.

*Nature* **604**, 21 (2022)

doi: <https://doi.org/10.1038/d41586-022-00776-2>

## References

1. Chornelia, A., Jianmei, L. & Hughes, A. C. *Front. Ecol. Evol.* **10**, 854509 (2022).
2. Zhou, P. *et al.* *Nature* **579**, 270–273 (2020).
3. Temmam, S. *et al.* *Nature* <https://doi.org/10.1038/s41586-022-04532-4> (2022).

---

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-022-00776-2>

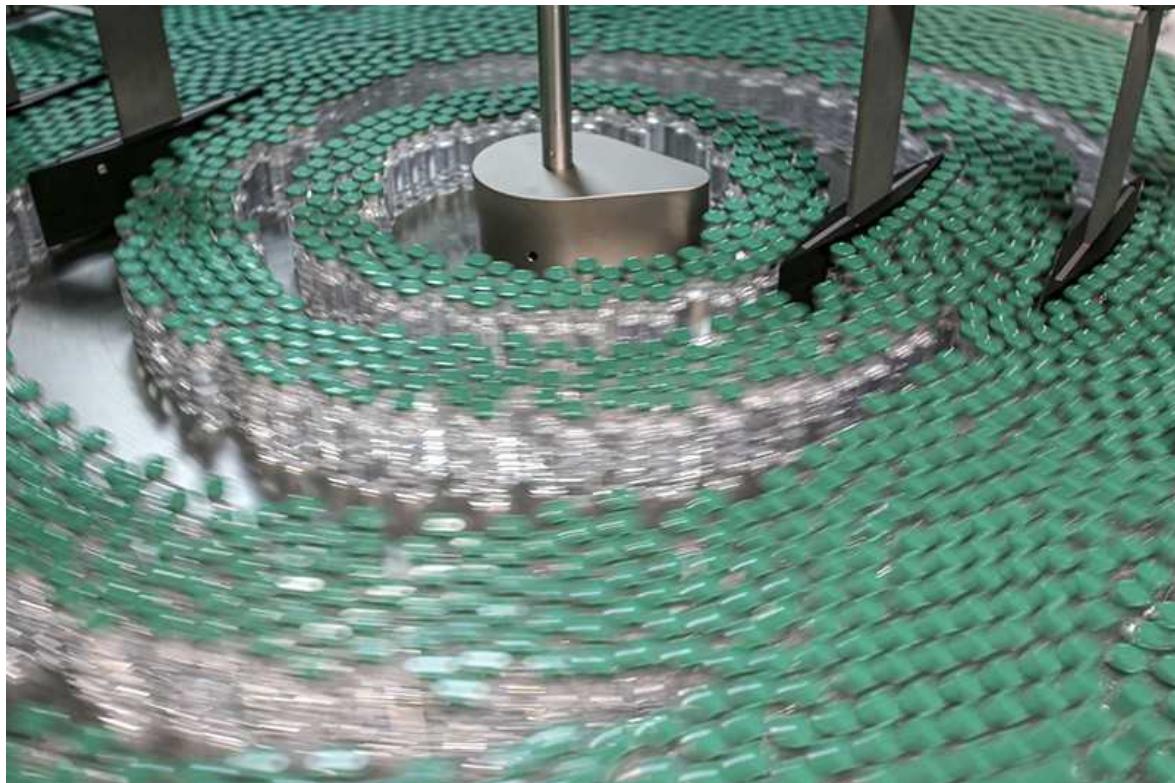
| [Section menu](#) | [Main menu](#) |

- NEWS FEATURE
- 05 April 2022

# Could computer models be the key to better COVID vaccines?

For vaccine dosing decisions, past experience and best guesses won the day in the mad rush to beat back the pandemic. Modelling tools might have made a difference.

- [Elie Dolgin](#) 0



Speed to approval and manufacturing capabilities factored in to dosing decisions for COVID-19 vaccines.Credit: Dhiraj Singh/Bloomberg via Getty

When Moderna joined the hunt for a coronavirus vaccine in early 2020, the company had only limited clinical experience with its technology.

Scientists had tested the company's messenger RNA (mRNA)-based vaccines against a few viruses, such as avian influenza and Zika, in humans. They found that the highest dose levels — upwards of 300 micrograms — often triggered undesirable side effects. The lowest doses (around 10 µg) did not always elicit a sufficient immune response.

There seemed to be a happy medium: in a two-dose vaccine for another respiratory virus with pandemic potential<sup>1</sup>, a new strain of bird flu, the sweet spot was around 100 µg. So, it made intuitive sense for Moderna, based in Cambridge, Massachusetts, and its collaborators at the US National Institute of Allergy and Infectious Diseases (NIAID) in Bethesda, Maryland, to try something similar to tackle SARS-CoV-2.

Within days of confirming that the coronavirus vaccine it had developed offered protection in mice, the company started human trials, testing doses of 100 µg to see whether its intuition was right. It also trialled 25- and 250-µg doses in case it wasn't.

“The whole point was to go quickly,” says Barney Graham, former deputy director of the NIAID’s Vaccine Research Center, who oversaw the vaccine’s early development. “That’s just how it’s done. It’s not a real precise process.”

Other efforts to develop COVID-19 vaccines followed a similar playbook. The University of Oxford, UK, in partnership with AstraZeneca in Cambridge, UK, started testing its vaccine — made from an engineered adenovirus — at a dose of 50 billion viral particles. It chose the dose in large part because that amount was used in previous trials of the same vaccine platform for other pathogens, including the coronavirus responsible for Middle East respiratory syndrome.

This approach to dose selection [produced several safe and effective COVID-19 vaccines in record time](#) — which has helped to save millions of lives around the world — but it did not necessarily take full advantage of the vaccines’ pandemic-altering potential, scientists say. Some blame

companies' inexact, educated guesses for a high rate of adverse events connected to many shots, the diminished efficacy of others and several high-profile trial failures, including initial attempts to develop a shot for young children.

"It still feels like people just threw a lot of spaghetti at the wall and saw what stuck," says Thomas Evans, chief scientific officer of Vaccitech in Oxford, which devised the adenovirus-mediated gene-delivery system found in the Oxford–AstraZeneca vaccine.

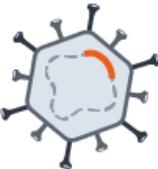
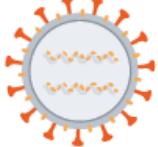
A growing number of scientists think that the industry can do better. With an eye towards optimizing immune responses, they have been developing mathematical and computational models over the past several years to inform dose decision-making for vaccine trials. Not everyone is convinced the models are ready for prime time; many aren't even aware that the platforms exist. But those who embrace the technology say that, if companies had simply capitalized on all the tools at their disposal, COVID-19 vaccines might be doing an even better job at containing viral spread and limiting collateral damage. "We missed a huge opportunity," Evans says.

## Leap of faith

Pharmaceutical companies have long used computational modelling strategies to fine-tune drug dosing, but such techniques have rarely been applied to vaccine development. Past experience and animal testing typically guide dose selection for experimental vaccines — and things were no different for those against COVID-19. The result was a range of doses for the products (see 'Dosing decisions').

# DOSING DECISIONS

Companies chose different dosing regimens for their vaccines, often for different reasons.

Vaccine type	Company	Dose*
<b>mRNA</b> 	Pfizer–BioNTech	30 µg
	Moderna	100 µg
<b>Adenovirus vectored</b> 	AstraZeneca–Oxford	50 billion viral particles
	Johnson & Johnson	50 billion viral particles
	Gamaleya (Sputnik V)	100 billion viral particles
<b>Inactivated virus</b> 	Sinopharm	4 µg
	Sinovac Biotech	3 µg
	Bharat Biotech	6 µg
<b>Protein**</b> 	Novavax	5 µg
	Biological E	25 µg

\*Approved adult dose for each shot

\*\*Antigen dose, not including adjuvant

©nature

Vaccine developers chose amounts that worked in other disease contexts, but immune responses can vary widely from one pathogen to the next. Animal studies add a degree of confidence in a vaccine's success — but the immune system of a mouse or monkey is not the same as that of a human, and scientists don't fully understand how to scale doses across species. So, most vaccine companies simply made what Jeff Barrett, a quantitative pharmacologist at the Critical Path Institute, a non-profit organization in

Tuscon, Arizona, describes as “a leap of faith” from animal models to human testing.

Moderna’s vaccine programme was no exception. Researchers involved in the earliest mouse studies administered two-shot regimens, with doses of up to 20 µg each<sup>2</sup>. But, according to Graham, they made little attempt to quantitatively map the immune responses observed in mice that received different doses to anticipated outcomes in people: the plan all along was to anchor human trials around the 100-µg dose that worked best for Moderna’s bird-flu vaccine candidate.

Company executives defend the approach because of the time and data constraints. “You make the best decision you can,” says Jacqueline Miller, head of infectious diseases at Moderna, “but that is informed by some of the previous programmes that had small phase I data with other vaccine antigens.”

Business considerations factor in as well. Pfizer in New York City and BioNTech in Mainz, Germany, opted for a shorter gap between doses for their mRNA jab, in part to help them beat Moderna in the race to marketing authorization, whereas Johnson & Johnson in New Brunswick, New Jersey, initially advanced a one-dose regimen to differentiate the company’s COVID-19 vaccine from others in development.

There were also public-health arguments for these dosing decisions, such as getting people vaccinated quickly. And the choices made during the sprint to provide a vaccine [had real-world consequences](#).

Moderna’s 100-µg shot proved to offer greater protection against infection, disease and hospitalization than the one from Pfizer–BioNTech, which used only 30 µg of mRNA per injection. By one estimate, recipients of the Pfizer–BioNTech vaccine had a 58% greater risk of infection from the Delta variant of SARS-CoV-2 than those who got Moderna’s shot<sup>3</sup>. (The trade-off is that Moderna’s product came with more frequent post-vaccine reactions.)

Differences in formulation and administration schedules could be at play, says John Moore, an immunologist at Weill Cornell Medicine in New York City. But he, like many researchers, points to dose size as the most probable

explanation for differences in efficacy and tolerability. Moderna's own head-to-head trials of 50- and 100- $\mu$ g dosing schemes support this conclusion<sup>4</sup>.

Perhaps a more consequential decision was the one that executives at Pfizer and BioNTech made for their child-sized COVID-19 vaccine.

In a small trial involving a few dozen children under 5 years old, the companies found that a pair of 3- $\mu$ g shots was sufficient to prompt an antibody response comparable to that in teenagers and young adults who had received two full doses. What's more, the mini-dose didn't trigger the severe fevers observed among children given a 10- $\mu$ g shot, so Pfizer and BioNTech moved ahead with the smallest dose.

But in a later-stage trial involving thousands of infants and young children, protection came up short. Two- to four-year-olds failed to produce enough antibodies, and a third booster might be necessary to develop adequate immune protection in those children.

Meanwhile, Moderna announced last month that a 25- $\mu$ g version of its shot provided the same level of immune protection against COVID-19 in children under six, as did a full 100- $\mu$ g dose in young adults. Moderna is now moving forwards with global regulatory submissions in every age bracket.

## Tricky business

With the benefit of hindsight, most scientists now think that Pfizer and BioNTech chose too low a dose for children under five. But it's hard to fault the thinking of drug executives, who were trying to minimize side effects, says Karim Azer, who has previously worked on tuberculosis-vaccine modelling at the Bill & Melinda Gates Medical Research Institute in Cambridge, Massachusetts. "The dose-response relationship with vaccines can be very tricky," he says.

There are several complicating factors. With conventional pharmaceuticals, greater drug concentrations usually yield more potent effects, at least up to a certain level. This isn't the case with vaccines, because higher doses can sometimes produce less favourable responses.



Dosing decisions might have slowed approval for a COVID vaccine in young children. Credit: Alexandre Schneider/Getty

That's because repeat exposure to vaccine antigens can cause certain arms of the immune system to secrete enough pro-inflammatory signalling molecules to trigger a phenomenon known as immune exhaustion, leading to impaired protection.

Timing is also important: a long interval between shots might coax out more protective antibodies, but one that's too long risks missing an optimal window. And dose-response dynamics often differ widely by age. A child is not simply a small adult when it comes to vaccines — more so than for other medicines.

Then, there's the question of what to measure when calculating vaccine-mediated protection. Antibodies or immune cells? Rates of infection or of disease and death?

“A definition of optimal dose may vary depending on which of these factors you care about,” says John Benest, a mathematical biologist at the London

School of Hygiene & Tropical Medicine (LSHTM).

Together with his supervisor Richard White and former group member Sophie Rhodes, Benest took published data from an early clinical study sponsored by CanSino Biologics in Tinjian, China, maker of a one-shot COVID-19 vaccine based on a viral vector. The researchers modelled<sup>5</sup> dosing schemes, prioritizing population-level immunity in one scenario, individual immunogenicity and safety in another and factoring in cost containment in a third. In every situation, the optimal dose — predicted on the basis of data from CanSino’s first-in-human trial — was more than double the amount now approved for use in China and elsewhere.

“It’s a shame,” says Rhodes, now a staff scientist at Certara, a drug-development consultancy headquartered in Princeton, New Jersey. Had she and other specialists in infectious-disease modelling been better positioned to obtain those kinds of quantitative insight early in the pandemic response, “it could have changed the way that we developed vaccines”, Rhodes says. CanSino chief scientific officer Tao Zhu stands by the company’s dosing decisions: the LSHTM analysis “is a good model”, he says, but it doesn’t account for logistical challenges in administering higher doses and later-stage trial data that shaped the company’s final call on which dose to use.

## Model makers

The LSHTM researchers, along with Evans and others, have been at the forefront of efforts to create a mathematical framework for informing vaccine dose decision-making<sup>6</sup>. In 2015, they convened the world’s first workshop dedicated to the topic. (Only a few dozen people attended.) In the years since, they have honed their modelling techniques with an eye to streamlining the process of determining vaccine doses for early-stage trials.

For any vaccine candidate, the modelling framework starts with data. The researchers feed immune-response results from animal experiments into their equations to produce a predicted dose–response curve. They then scale that dose–response relationship to humans using clinical data from a more limited number of doses, often from historical work on similar vaccines. In this way, they come up with expected ‘best’ doses for testing in human trials

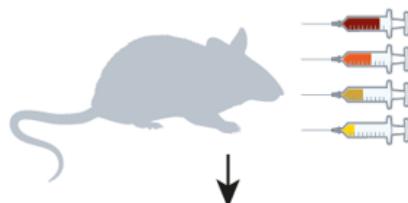
— and they can further refine the model's predictions as more data become available (see 'Immune modelling').

# IMMUNE MODELLING

Scientists have been developing mathematical models to optimize vaccine doses before they are given to humans.

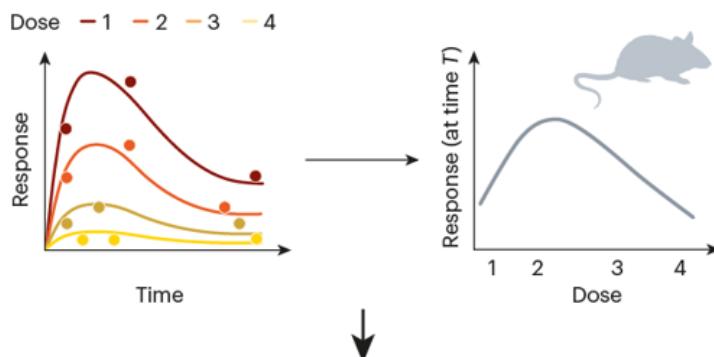
## 1. Animal testing

Scientists administer a wide range of doses in small animal models.



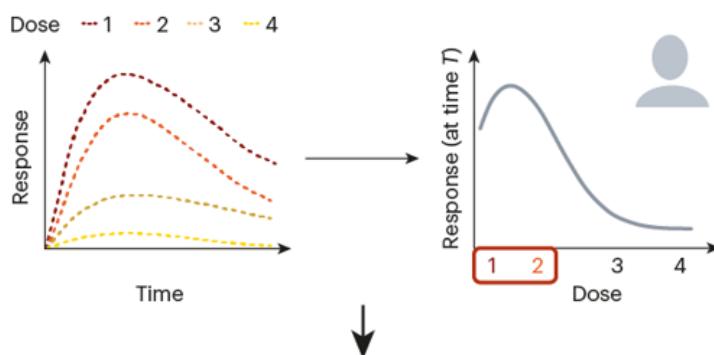
## 2. Initial modelling

Researchers identify doses correlating to the largest and smallest immune responses in animals. They use modelling to estimate the relationship between dose and response.



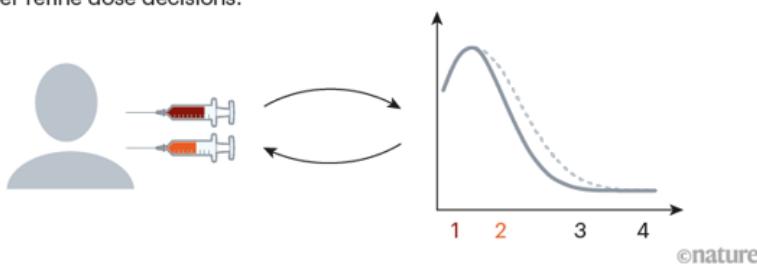
## 3. Translation to humans

Data from the animal model are scaled up to predict a theoretical dose-response relationship for humans, which helps scientists to choose initial doses for testing.



## 4. First-in-human

Scientists feed data from early human testing into the human-response model to further refine dose decisions.



Source: Adapted from ref. 6

As a proof of principle, the researchers fit their model with mouse and human data on [the response of T cells](#) — a type of immune cell — to an experimental tuberculosis vaccine. The mathematics then predicted that lower doses would offer the best immune response<sup>7</sup>. Independent clinical studies run in parallel to the group's modelling project bore this out.

Jennifer Linderman, a systems biologist at the University of Michigan in Ann Arbor, says such approaches could be useful for guiding dose decisions in the future. “We’re in a position now where, going forward, we can be much more intentional about vaccine design,” she says. She and Denise Kirschner, a computational immunologist at the University of Michigan Medical School, developed HostSim, a model that incorporates lung biology alongside simulations of what happens in the blood and lymph nodes<sup>8</sup>.

Although initially focused on tuberculosis, Kirschner notes that, with the appropriate data inputs, her team’s tool could guide vaccine development for any pathogen that infects the airways. “We can use our model for flu. We can use our model for COVID. We can use it for lots of things,” she says.

In simulated trials, at least, such modelling approaches allow vaccine developers to vet many more doses than would otherwise be feasible, even in the largest of clinical studies. “You can test a wider range of scenarios *in silico*,” says Luca Marchetti, a computer scientist at the Microsoft Research–University of Trento Centre for Computational and Systems Biology in Rovereto, Italy, who last year developed a model to support mRNA-vaccine development<sup>9</sup>.

## Pandemic response

Before the pandemic, few companies wanted to invest in this kind of vaccine-dose modelling. Around five years ago, Evans and Kent Kester, then head of translational sciences at Sanofi Pasteur (the vaccines division of Paris-based Sanofi), tried to create a research consortium focused on tool development in this area, but they failed to get buy-in from industry or regulatory authorities. “No one was really interested in pursuing it,” says

Kester, now vice-president of translational medicine at IAVI, a vaccine non-profit organization based in New York City.

Because of COVID-19, more vaccine manufacturers are now experimenting with dose modelling, and regulatory agencies are monitoring the science closely.

“This is an important area,” says Marco Cavaleri, head of biological health threats and vaccines strategy at the European Medicines Agency in Amsterdam. “The more we can refine these techniques, the more we will be prepared in the future.”

Last June, the US Food and Drug Administration convened a workshop at which researchers discussed best practices for modelling vaccine dose-response relationships. White and Rhodes spoke, as did Andrzej Kierzek and Piet van der Graaf, modellers at Certara who, before the pandemic, had created a tool for running virtual trial simulations of antibody drugs and biological therapies.

The tool had helped pharmaceutical companies to predict unwanted immune reactions. But when COVID-19 hit, the Certara researchers realized that the same model could be used to forecast desired immune responses from vaccines. As a first test, they plugged in the amino-acid sequence corresponding to the coronavirus’s spike protein, the bit used by most COVID-19 vaccines. As van der Graaf recalls: “We got a surprising, meaningful result.”

The immune responses predicted by the model “seemed to be plausible”, he says. And as companies such as Moderna and Pfizer–BioNTech began to publish more human and mouse data, the Certara scientists would incorporate those results into their simulation workflow. They added response dynamics for T cells and B cells, which produce antibodies, into the mix, along with plug-in modules to account for different vaccine technologies and routes of administration.

Over time, their model — dubbed the Vaccine Simulator — grew in sophistication. And before efficacy results were even known for the first wave of COVID-19 vaccines, Kierzek and van der Graaf had already

concluded that longer dosing intervals than those being evaluated would yield improved antibody responses<sup>10</sup>. Data from the United Kingdom, where extended dosing schedules were routine during the vaccine roll-out, later confirmed that advantage.

## Dose decisions

Daiichi Sankyo was one of the first drug companies to incorporate the Certara platform into its vaccine-development programme. The Tokyo-based firm began testing its mRNA vaccine in humans in March 2021 — a slow start that gave trial organizers the opportunity to learn from the experiences of other companies.

Scientists at Daiichi looked at the dosing and scheduling regimens of other mRNA shots, and determined that a dose between the 30 µg used by Pfizer–BioNTech and the 100 µg used by Moderna might provide the ideal balance of immunogenicity and tolerability. They planned to press forwards with an initial trial evaluating up to 60-µg doses of mRNA.

But when the company, in partnership with Certara, simulated immune responses to the vaccine in virtual participants, they found that older individuals failed to mount robust antibody responses at the highest planned dose. “This result triggered an internal discussion on the phase I study design,” says Daiichi’s Ryoko Sawamura, who leads a modelling team at the company. Ultimately, her firm added a 100-µg dose to its first-in-human trial protocol.

AstraZeneca scientists are now using Certara’s model to simulate scenarios not captured by earlier clinical studies of the company’s vaccine. They are interrogating immune responses in populations that were under-represented in trials — particular ethnic groups, for example, and immunocompromised people — to predict who might benefit from non-standard dosing. And they are looking at long-term immunity trends to inform optimal timing of booster-dose regimens.

Few seasoned vaccine developers are converts to the approach. “There are too many variables to model in a way,” says Emilio Emini, chief executive

of the Gates Medical Research Institute and a former vaccine research head at Pfizer and Merck. “At the moment, there are no clear prospective models that exist that allow one to make that initial prediction — other than extrapolating as best as one can,” he adds.

But such modelling tactics are catching on. The FDA says that, at the end of 2021, it received its first submission of a vaccine product created using modelling to optimize dose–response relationships.

Although Moderna’s Miller and other industry executives say it’s too soon to begin prospectively selecting vaccine doses for human trials, they might change their minds if and when the tools get validated and prove their worth. “As we gain more experience,” Miller says, “we’ll get there.”

*Nature* **604**, 22–25 (2022)

*doi:* <https://doi.org/10.1038/d41586-022-00924-8>

## References

1. Feldman, R. A. *et al.* *Vaccine* **37**, 3326–3334 (2019).
2. Corbett, K. S. *et al.* *Nature* **586**, 567–571 (2020).
3. Dickerman, B. A. *et al.* *N. Engl. J. Med.* **386**, 105–115 (2022).
4. Chu, L. *et al.* *Vaccine* **39**, 2791–2799 (2021).
5. Benest, J., Rhodes, S., Quaife., M., Evans, T. G. & White, R. G. *Vaccines* **9**, 78 (2021).
6. Rhodes, S. J., Knight, G. M., Kirschner, D. E., White, R. G. & Evans, T. G. *J. Theor. Biol.* **465**, 51–55 (2019).
7. Rhodes, S. J. *et al.* *NPJVaccines* **3**, 36 (2018).
8. Joslyn, L. R., Linderman, J. J. & Kirschner, D. E. *J. Theor. Biol.* **539**, 111042 (2022).

9. Sevaggio, G. *et al.* *CPT Pharmacometrics Syst. Pharmacol.* **10**, 1448–1451 (2021).
  10. Giorgi, M., Desikan, R., van der Graaf, P. H. & Kierzek, A. M. *CPT Pharmacometrics Syst. Pharmacol.* **10**, 1130–1133 (2021).
- 

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-022-00924-8>

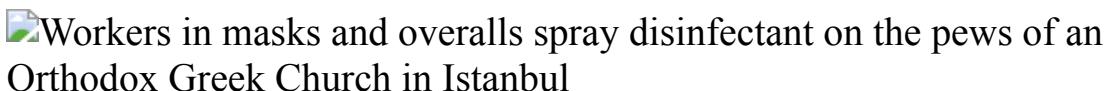
| [Section menu](#) | [Main menu](#) |

- NEWS FEATURE
- 06 April 2022

# Why the WHO took two years to say COVID is airborne

Early in the pandemic, the World Health Organization stated that SARS-CoV-2 was not transmitted through the air. That mistake and the prolonged process of correcting it sowed confusion and raises questions about what will happen in the next pandemic.

- [Dyani Lewis](#) <sup>0</sup>



Public-health advice on COVID-19 in early 2020 focused on sanitizing surfaces more than protecting against airborne transmission. Credit: Ozan Kose/AFP/Getty

As 2021 drew to a close, the highly contagious Omicron variant of the pandemic virus was racing around the globe, forcing governments to take drastic actions once again. The Netherlands ordered most businesses to close on 19 December, Ireland set curfews and many countries imposed travel bans in the hope of taming the tsunami of COVID-19 cases filling hospitals. Amid the wave of desperate news around the year-end holidays, one group of researchers hailed a development that had seemed as though it might never arrive. On 23 December, the World Health Organization (WHO) uttered the one word it had previously seemed incapable of applying to the virus SARS-CoV-2: ‘airborne’.

On its website, a page titled ‘Coronavirus disease (COVID-19): How is it transmitted?’ was quietly edited to state that a person can be infected “when infectious particles that pass through the air are inhaled at short range”, a

process otherwise known as “short-range aerosol or short-range airborne transmission”. The website says that transmission can occur through “long-range airborne transmission” in poorly ventilated or crowded indoor settings “because aerosols can remain suspended in the air or travel farther than conversational distance”.

“It was a relief to see them finally use the word ‘airborne’, and to say clearly that airborne transmission and aerosol transmission are synonyms,” says aerosol chemist Jose-Luis Jimenez at the University of Colorado Boulder.

The seemingly uncontroversial statement marked a clear shift for the Switzerland-based WHO, which had tweeted categorically early in the pandemic, [“FACT: #COVID19 is NOT airborne.”](#), casting the negative in capital letters as if to remove any doubt. At that time, the agency maintained that the [virus spreads mainly through droplets](#) produced when a person coughs, sneezes or speaks, an assumption based on decades-old infection-control teachings about how respiratory viruses generally pass from one person to another. The guidance recommended distancing of more than one metre — within which these droplets were thought to fall to the ground — along with [hand washing and surface disinfection](#) to stop transfer of droplets to the eyes, nose and mouth.

It took until 20 October 2020 for the agency to acknowledge that aerosols — tiny specks of fluid — can transmit the virus, but the WHO said this was a concern only in specific settings, such as [indoor, crowded and inadequately ventilated spaces](#). Over the next six months, the agency gradually altered its advice to say that aerosols could carry the virus for more than a metre and remain in the air (see ‘[Changing views of how COVID spreads](#)’).

But this latest tweak is the WHO’s clearest statement yet about airborne transmission of SARS-CoV-2. And it places the virus among a select group of ‘airborne’ infections, a label long reserved for just a handful of the world’s most virulent pathogens, including measles, chickenpox and tuberculosis.

The change brings the WHO’s messaging in line with what a chorus of aerosol and public-health experts have been trying to get it to say since the earliest days of the outbreak. Many decry the agency’s slowness in stating

— unambiguously — that SARS-CoV-2 is airborne. Interviews conducted by *Nature* with dozens of specialists on disease transmission suggest that the WHO’s reluctance to accept and communicate evidence for airborne transmission was based on a series of problematic assumptions about how respiratory viruses spread.

For example, even in the middle of the fast-moving epidemic, the WHO dismissed field epidemiology reports as proof of airborne transmission because the evidence was not definitive, something that is difficult to achieve quickly during an outbreak. Other criticisms are that the WHO relies on a narrow band of experts, many of whom haven’t studied airborne transmission, and that it eschews a precautionary approach that could have protected countless people in the early stages of the pandemic.

Critics say that inaction at the agency led to national and local health agencies around the world being similarly sluggish in addressing the airborne threat. Having shifted its position incrementally over the past two years, the WHO also failed to adequately communicate its changing position, they say. As a result, it didn’t emphasize early enough and clearly enough the importance of ventilation and indoor masking, key measures that can prevent airborne spread of the virus. Lidia Morawska, an aerosol scientist at the Queensland University of Technology in Brisbane, Australia, spearheaded several efforts to convince the WHO and other health agencies of the airborne threat. She says that airborne transmission was “so obvious” as far back as February 2020, and that omitting it from official guidelines was disastrous.

But Dale Fisher, an infectious-diseases physician at the National University Hospital in Singapore and chair of the WHO’s Global Outbreak Alert and Response Network steering committee, doesn’t think that confusion over whether the virus is airborne has had a defining impact on how the pandemic has played out. “It’s not the cause of the catastrophe we’ve seen,” he says.

Some other researchers defend the agency’s response, given the rapidly evolving situation. “I really don’t think anybody dropped the ball, including WHO,” says Mitchell Schwaber, an infectious-diseases physician at Israel’s ministry of health and an external adviser to the WHO. “So many

assumptions that we had about this virus were proven false. We always, we always were learning new things.”

Resolving this debate about how to assess the transmission of respiratory viruses matters, say researchers, because a more deadly variant of SARS-CoV-2 could emerge at any time, and new respiratory viruses will almost certainly plague humanity at some point. It’s not clear whether the WHO and the world will be ready.

## Tension in the air

In the final days of March 2020, Morawska contacted dozens of colleagues — an international mix of aerosol scientists, infectious-disease specialists, and building and ventilation engineers — to get the word out about the airborne threat of SARS-CoV-2. On 1 April 2020, the group sent an e-mail laying out their case to Michael Ryan, head of the WHO’s Health Emergencies Programme, and Maria Van Kerkhove, technical lead of the WHO’s COVID-19 response.

Within an hour, the agency was on the phone. Two days later, the group attended a video conference with members of the Health Emergencies Programme and the Infection Prevention and Control Guidance Development Group (IPC GDG) — an external group of about 40 clinicians and researchers that advises the WHO on infection containment, especially in hospitals. At the time of the meeting, more than one million people had been infected with SARS-CoV-2, and 54,000 had died. Community spread was rampant in several countries.

Morawska presented what she says was a compelling case for airborne transmission. Two facts stood out. First, there was solid evidence that people were becoming infected even when they were more than one metre — the safe distance recommended by the WHO — from a contagious individual. Second, years of mechanistic studies had demonstrated how mucus in a person’s airway can spray into aerosols during speech and accumulate in stagnant rooms. Morawska felt rebuffed by the WHO and its advisers. “I didn’t have a feeling that they were trying to see this from our perspective,” she says.

She and other people who study aerosols and airborne disease transmission say that the IPC GDG is ill-equipped to assess this type of transmission because most of its members have focused on controlling infections in hospitals and they lack expertise in the physics of how airborne contagions spread. At the time of the 1 April meeting, no one in the IPC GDG had studied this type of disease transmission, say critics.

“If it is a new disease, you better include everyone,” says Yuguo Li, a building environment engineer at the University of Hong Kong, whose study of the SARS outbreak in 2002–03 had concluded that the virus responsible, SARS-CoV, probably spread through the airborne route<sup>1</sup>. He suspected that SARS-CoV-2 was also airborne, although he initially thought that only short-range airborne transmission was likely.

Marcel Loomans, an indoor-air-quality physicist at Eindhoven University of Technology in the Netherlands, says that it is often hard to find common ground between the two disciplines. “On the medical side, they were not aware of how aerosols behave in the air and what ventilation can do,” he says. People end up “talking past each other”.



Early WHO advice on masks recommended them only for infected people and their carers.Credit: S.C. Leung/SOPA Images/LightRocket/Getty

The disconnect was there even in the use of scientific terms. Infection-control experts have long drawn a hard line between droplet viruses and airborne ones, seeing only the latter as capable of travelling far and lingering in the air. “Dogmatic bias is certainly a big part of it,” says Don Milton, an occupational-health physician who studies aerosol transmission of infectious diseases at the University of Maryland in College Park. He says that he was disappointed but not surprised by the WHO’s lack of action in addressing the airborne threat after the 1 April meeting. “I’m just familiar with how the medical profession thinks,” he says.

But Schwaber, who chairs the IPC GDG, recalls the meeting differently. “We took very seriously the issues that they raised at the meeting, and responded to them,” he says. “Nothing was being blown off, nothing was being ignored.”

At the time, he says, the available evidence suggested that airborne precautions throughout hospitals — including N95 masks for staff, visitors and patients — were unnecessary. Still, faced with soaring deaths among frontline doctors and nurses, most hospitals and health agencies adopted these precautions on their COVID-19 wards, as well as less-stringent protections such as wearing surgical masks in other areas of the hospital.

Mark Sobsey, an environmental microbiologist at the University of North Carolina in Chapel Hill who is a member of the IPC GDG, says that especially in the early days, the concerns brought to the WHO about airborne transmission were “largely unfounded” and lacked credible evidence, such as the isolation of infectious virus particles from air samples. Epidemiological data from outbreak investigations were “especially weak”, he says.

According to Trish Greenhalgh, a primary-care health researcher at the University of Oxford, UK, the IPC GDG members were guided by their medical training and the dominant thinking in the medical field about how infectious respiratory diseases spread; this turned out to be flawed in the case of SARS-CoV-2 and could be inaccurate for other viruses as well.

These biases led the group to discount relevant information — from laboratory-based aerosol studies and outbreak reports, for instance. So the IPC GDG concluded that airborne transmission was rare or unlikely outside a small set of aerosol-generating medical procedures, such as inserting a breathing tube into a patient.

That viewpoint is clear in a commentary by members of the IPC GDG, including Schwaber, Sobsey and Fisher, published in August 2020<sup>2</sup>. The authors dismissed research using air-flow modelling, case reports describing possible airborne transmission and summaries of evidence for airborne transmission, labelling such reports “opinion pieces”. Instead, they concluded that “SARS-CoV-2 is not spread by the airborne route to any significant extent”.

In effect, the group failed to look at the whole picture that was emerging, says Greenhalgh. “You’ve got to explain all the data, not just the data that you’ve picked to support your view,” and the airborne hypothesis is the best fit for all the data available, she says. One example she cites is the propensity for the virus to transmit in ‘superspread events’, in which numerous individuals are infected at a single gathering, often by a single person. “Nothing explains some of these superspread events except aerosol spread,” says Greenhalgh.

Throughout 2020, there was also mounting evidence that indoor spaces posed a much greater risk of infection than outdoor environments did. An analysis of reported outbreaks recorded up to the middle of August 2020 revealed that people were more than 18 times as likely to be infected indoors as outdoors<sup>3</sup>. If heavy droplets or dirty hands had been the main vehicles for transmitting the virus, such a strong discrepancy would not have been observed.

Although the WHO played down the risk of airborne transmission, it did invite Li to become a member of the IPC GDG after he spoke to the group in mid-2020. Had the organization not at least been open to his view that infections were caused by aerosols, especially at short range, “they would not have invited me there as they knew my standing”, he says.

Still, Li is disappointed that it took the WHO until October 2020 to acknowledge that aerosols play a part in disease transmission in community settings. And in its updated guidelines on mask use, in December 2020, the agency still emphasized shortfalls and gaps in the evidence for aerosol transmission, and the need for more “high quality research” to understand the specifics of how the virus spreads. It wasn’t until the end of April 2021 that long-range aerosol transmission was added to a question-and-answer section on the agency’s website about how the virus spreads. And the term airborne wasn’t officially added until December 2021.

## Conservative approach

Some scientists note that the WHO’s decision to classify SARS-CoV-2 as airborne, belated as it was, is momentous. That’s because it flies in the face of the established view of respiratory virus transmission that held sway when the pandemic began — that nearly all infectious diseases are spread by droplets, not through the air. And researchers say that this change is particularly important because the organization generally takes a conservative approach. “What the WHO says is normally based on a consensus of expert advice and opinion,” says Christopher Dye, an epidemiologist who served as the scientific adviser to the agency’s director-general until 2018.

And although the WHO has drawn strong criticism for the way in which it assessed SARS-CoV-2 transmission, some researchers don’t find the agency’s response surprising. The international community looks to the WHO for early warnings of disease outbreaks. But when it comes to science, the agency “sees its role as certifying the current expert consensus, not (usually) advancing new, tentative knowledge”, says Peter Sandman, an independent risk-communications specialist based in New Jersey who has worked as a consultant to the WHO.

Schwaber says: “Individuals and governments and public-health bodies are looking to a WHO GDG, not to conjecture. They’re looking to a WHO GDG to put out guidance. That everything that we say can be backed by evidence.”

The WHO frequently gets attacked, “so you can understand how they’d be risk averse”, says Tom Frieden, president of the global-health initiative Resolve to Save Lives and former head of the US Centers for Disease Control and Prevention (CDC). Frieden is critical of some aspects of the WHO’s pandemic response, including how slow it was to recommend the use of masks. But he says that the agency is in a difficult position during health crises.

In 2009, for instance, it was accused of being alarmist over the H1N1 swine influenza outbreak that petered out with few lives lost. “WHO got hit hard for that,” says Dye, even though he thinks the agency was right to be cautious and declare a public-health emergency of international concern.

## Hard line to tread

Virologist May Chu, a member of the IPC GDG at the Colorado School of Public Health in Aurora, says that the WHO treads a difficult line, and tends to be quite conservative in its recommendations to avoid putting out information that later proves to be incorrect. “You can’t be backtracking” on advice, adds Fisher, because “then you lose complete credibility”.

The gravity of the situation might have made the WHO even more cautious in its pronouncements and less likely to stray from consensus views, according to Sandman’s partner Jody Lanard, an independent risk-communications specialist who has also worked with the WHO in the past.

In previous situations — such as during the Ebola outbreak in West Africa, and in polio vaccine campaigns — the WHO was more nimble than it has been during the COVID-19 pandemic, Lanard says. “I’ve seen them be able to change what their approach was, or try different things,” she says. But during the pandemic “it’s so tempting to be very, very cautious”, because millions of lives will be affected by the agency’s recommendations.

Loomans and others question why, when concerns were growing that SARS-CoV-2 could be airborne, the WHO didn’t adopt a precautionary approach by acknowledging the possibility of different risks, even without definitive proof.



Schoolchildren in Taipei eat lunch behind partitions to stop the spread of COVID-19 in April 2020, after the WHO stressed the dangers of respiratory droplets that travel short distances. Credit: Sam Yeh/AFP/Getty

And in May 2021, the Independent Panel for Pandemic Preparedness and Response (IPPPR), a body established by the WHO a year earlier to review the agency's actions at the start of the pandemic, called out the WHO for not applying the precautionary principle to another crucial aspect of COVID-19 transmission — whether it could spread from human to human (see [go.nature.com/3iqhfjm](https://go.nature.com/3iqhfjm)). “There is a case for applying the precautionary principle in any outbreak caused by a new pathogen resulting in respiratory infections, and thereby for assuming that human-to-human transmission will occur unless the evidence specifically indicates otherwise,” the IPPPR said in its 2021 report.

In practice, applying the precautionary approach to the question of how SARS-CoV-2 — or any newly emerged pathogen — is transmitted would mean initially assuming that all routes of transmission are possible. “That should be your starting point, and then you can strike out routes if you’re sure,” says Loomans.

But Schwaber says that this approach carries risks. “To say, well, the best interests of the patient and the best interests of the health-care worker involve invoking the precautionary principle would also imply that there’s no downside to invoking it,” he says. Taking full precautions against airborne transmission would require major changes at hospitals, such as using negative-air-pressure isolation rooms and uncomfortable N95 masks for all staff and visitors. Such changes need to be weighed against the evidence that they are required, he says.

Sobsey says that the WHO did adopt the precautionary principle, in part because of the advice from aerosol scientists. That’s why, he says, the agency stated in July 2020 that airborne transmission couldn’t be ruled out — and why it started placing more emphasis on ventilation as a protective measure, even though the evidence for airborne transmission was weak at the time.

“They are not totally wrong,” says Li of those who claimed there were gaps in the evidence for airborne transmission, especially over larger distances. “It’s nothing bad to seek solid scientific evidence,” he says, but “when you see the spread so significantly, do you still wait for a nice *Nature* or *Science* article?” he says.

Still, other health organizations moved faster than the WHO despite the uncertainty. In February 2020, Li was contacted by the Chinese Center for Disease Control and Prevention for advice on air conditioning in public buildings and on public transport. At Li’s suggestion, he says, the centre recommended maximizing airflow in buildings from the outside, to help flush out any airborne contagion. At the time, Li didn’t think that ventilation would substantially reduce infection from a virus that he suspected was airborne only over short distances — an assumption that he later disproved. But he recommended improved ventilation because “I always support a precautionary approach,” he says.

## Communication problems

One thing that’s still missing, says Jimenez, is a clear communication campaign from the WHO. Its director-general, Tedros Adhanom

Ghebreyesus, acknowledged the challenges in his opening remarks at the agency's global conference on communicating science during health emergencies, on 7 June 2021. "Scientific processes, decision-making in an emergency context and mass communication do not fit together easily," Tedros said, adding that "high-quality research takes time, but time is something we don't have in an emergency".

During the early months of the pandemic, the WHO was fighting battles on other fronts. While it grappled with shortages of protective equipment and ventilators, it was also contending with misinformation about unproven treatments for COVID-19 and US threats to pull its funding from the organization.

But critics say that even two years into the pandemic, the WHO hasn't clearly communicated the risks from airborne transmission. And, perhaps as a result, governments around the world spent much of the pandemic focusing on hand washing and surface cleaning, instead of ventilation and indoor masking.

"The cacophony of changing messages has undoubtedly contributed greatly to resistance to masks and other measures," says Jimenez.

On 15 December 2021, less than two weeks before the latest change in wording on the WHO's website, Jimenez [put out a call on Twitter](#) for evidence of how governments and organizations either "don't know how to protect their citizens, or use @WHO's ambiguity to avoid doing so". He enumerated more than 100 examples in which health advice at the time was at odds with airborne precautions, indicating that the message was not filtering out from the agency.

Jimenez has continued to receive such examples. Now that the agency has changed the wording on its main website, Jimenez can call out these 'COVID Hall of Shame' offenders, as he labels them, for providing advice that is no longer in line with the international health agency.

"That is the arrogance, a bit, of what WHO is," says Chu. "Once you post [new guidance], it's pretty passive. They expect you to come to their website. They don't necessarily broadcast it."

But that's exactly what's needed, says Jimenez, especially given early communications that still haunt the agency, such as its tweet about COVID-19 not being airborne. "No doubt we owe the persistence of misinformation to that WHO announcement and firm position, at the time in which we were all scared and eager to learn how to protect ourselves, very early in the pandemic," says Jimenez.

The agency defends its actions throughout the pandemic. In a statement to *Nature* last month, a spokesperson said: "WHO has sought the expertise of engineers, architects and aerobiologists along with expertise in infectious diseases, infection prevention and control, virology, pneumology and other fields since the early days of the COVID-19 pandemic. In August 2020, we established the Environment and Engineering Control Expert Advisory Panel (ECAP) for COVID-19 to provide expert contributions for the development of guidance through evaluation and critical interpretation of available evidence (benefits and harm of interventions) related to relevant technical questions including indoor air quality management and ventilation as an engineering control measure in the context of COVID-19."

The organization says that initial guidance covered airborne precautions in health-care settings, but notes that: "As the evidence on the transmission of COVID-19 has expanded, we have learnt that smaller-sized infectious particles known as aerosols also play a role in transmission in community settings, and WHO has adapted its guidance and messages to reflect this in the December 2020 update to our mask guidance."

In response to critics who say that it hasn't adequately highlighted the changes it has made regarding the risks of airborne transmission, the WHO says that it has held about 250 press briefings and hundreds of live social-media events during the pandemic. It adds that it also pushes out information through social-media channels, meetings with doctors and mailing lists to scientists.

That's not enough, according to some researchers. Stephanie Dancer, a microbiologist at the Edinburgh Napier University, UK, says that the WHO needs to be clear about its position so that others follow its lead. "They have to show true strength of character and stand up and say, 'We got it wrong.'

We're going to get this right. Here are our next set of guidelines. This is where we're going to go. This is what we advise,"" she says.

## Off to a bad start

Part of the problem was how emphatic the WHO was at the beginning of the pandemic, says Heidi Tworek, a historian and public-policy specialist at the University of British Columbia in Vancouver. "To say that COVID was definitively not airborne unfortunately meant there was a massive hill to climb to undo that," she says. Right from the beginning, the WHO and other public-health authorities and governments should have emphasized that SARS-CoV-2 was a new coronavirus, and that guidelines would inevitably change, she says. "And when they do, it's a good thing because it means we know more."

"We're really talking here about two failures, not one," says Sandman. "Being reluctant to change your mind, and being reluctant to tell people you changed your mind." Like other public-health and scientific organizations, the WHO "are afraid of losing credibility by acknowledging that they got something wrong", he says.

But when Lanard worked with the WHO in 2005 to draft its risk-communications guidelines, one tenet that she advocated — to admit mistakes and errors when they occur — was removed from the final draft. She says that there were good reasons behind that decision, including that health officials in some countries could have faced imprisonment — or worse — if they had promoted information from the WHO that turned out to be incorrect. Officials and scientific advisers in several countries have [received death threats during the pandemic](#). "Inevitably you'll get it wrong sometimes," says Frieden. And the WHO is in a position that means "whatever they do, they get attacked", he says.

On the science front, questions remain about how much of COVID-19 transmission is airborne. Sobsey says that researchers still need to come up with evidence that the airborne route makes "an important contribution to the overall disease burden". Many on the other side of the aisle, such as Jimenez, are convinced that airborne transmission predominates. The US

Office of Science and Technology Policy voiced strong support for this view on 23 March, when its head, Alondra Nelson, issued a statement called '[Let's Clear the Air on COVID](#)', which said "the most common way COVID-19 is transmitted from one person to another is through tiny airborne particles of the virus hanging in indoor air for minutes or hours after an infected person has been there."

Other viruses long suspected of being airborne — including influenza and common cold viruses — will also be scrutinized. In September 2021, the US National Institutes of Health awarded Milton a multimillion-dollar grant to conduct trials that will determine whether airborne or droplet routes lead to influenza infection.

Li says that there's much greater recognition of airborne transmission because of the COVID-19 pandemic, and research over the next few years will probably show that most respiratory viruses can spread in this way. So the whole world will be more alert to the possibility of the airborne threat when old or new infectious diseases start spreading.

In the WHO, too, attitudes have shifted, according to Sobsey. "I think there's been a sea change in thinking at WHO as a consequence of the experience with this virus," he says, "which is — be more precautionary, even if you're not sure."

*Nature* **604**, 26–31 (2022)

*doi:* <https://doi.org/10.1038/d41586-022-00925-7>

## References

1. Yu, I. T. S., Wong, T. W., Chiu, Y. L., Lee, N. & Li, Y. *Clin. Infect. Dis.* **40**, 1237–1243 (2005).
2. Conly, J. *et al. Antimicrob. Resist. Infect. Control* **9**, 126 (2020).
3. Bulfone, T. C., Malekinejad, M., Rutherford, G. W. & Razani, N. *J. Infect. Dis.* **223**, 550–561 (2021).

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-022-00925-7>

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

## Books & Arts

- **[Global population is crashing, soaring and moving](#)** [ 04 April 2022]  
Book Review • From Japan to Yemen, India to Ukraine, rates of births, deaths and displacement are reshaping nations.
- **[Taming stars, and how do birds smell? Books in brief](#)** [ 11 March 2022]  
Book Review • Andrew Robinson reviews five of the week's best science picks.

- BOOK REVIEW
- 04 April 2022
- Correction [07 April 2022](#)

# Global population is crashing, soaring and moving

From Japan to Yemen, India to Ukraine, rates of births, deaths and displacement are reshaping nations.

- [Josie Glausiusz](#) <sup>0</sup>



Schoolchildren attend an open-air class in Yemen.Credit: Ahmad Al-Basha/AFP/Getty

## **8 Billion and Counting: How Sex, Death, and Migration Shape Our World** Jennifer D. Sciubba W. W. Norton & Company (2022)

In January 2022, the National Bureau of Statistics of China made a startling announcement: for the fifth year in a row, the country's birth rate had fallen. In 2021, 10.6 million children were born in China, compared with 12 million the year before. This was despite the government repealing its onerous one-child policy, permitting couples to have up to three children as of 2021.

As it turns out, many women in China feel that they cannot afford to have more than one child — or any. As death rates approach birth rates, China's population might soon begin to shrink. Likewise, "Japan is ageing so rapidly that if current trends continue, the nation could eventually disappear altogether", writes Jennifer Sciubba in her data-packed book *8 Billion and Counting*.

Almost eight billion people live on Earth; their futures are highly divergent, argues Sciubba, a senior associate at Washington DC think tank the Center for Strategic and International Studies. The twenty-first century "is less a story about exponential population growth than it is a story about differential growth — marked by a stark divide between the world's richest and poorest countries", she writes.

In Latin America and the Caribbean, eastern and southeast Asia, Europe and North America, Australia and New Zealand, the total fertility rate (TFR), or average number of children a woman is likely to have in her lifetime, was below replacement level (around 2.1 children per woman) in 2020. By contrast, sub-Saharan Africa's population is set to increase sixfold this century; its TFR is 4.72, down from 5.88 two decades ago. In Nigeria, children and adolescents are half of the population. In rural towns in South Korea, primary schools are closing for lack of pupils, whereas urban areas of Lagos ring "with the sounds of children playing", writes Sciubba. By 2050, Ethiopia and the Democratic Republic of the Congo (DRC) will be among the world's top ten most populous countries, joining long-standing members such as Brazil and India.

In some regions, population pressures are exacerbating problems such as poor governance, civil war and environmental devastation. According to the

annual Fragile States Index, produced by the Fund for Peace in Washington DC, in most of the countries in imminent danger of collapse — including South Sudan, Somalia and the DRC — more than half the population is under the age of 20 and fertility is high. In Somalia, young people can rarely find meaningful jobs. In Yemen, 3,000 children are born into poverty every day, Sciubba relates.

In countries with low fertility and an ageing population, social systems strain to support older people. In immigration-averse Japan, for example, robots including furry “carerobo” animals are assisting the dwindling number of elder-care workers. A 2021 report projects that by 2045, one-quarter of Japanese people aged older than 65 will have dementia ([N. Nakahori et al. \*BMC Geriatr.\* 21, 602; 2021](#)). As more people exit the labour force than enter it, wealthy countries worry about how to preserve economic growth. To soften the economic blow of ageing, Sciubba presents four options: raise retirement ages, cut benefits, push more people in the country to work or increase immigration.

She focuses at length on immigration. Only 2–4% of the world’s people live outside their country of origin, but between 2000 and 2020, according to the United Nations, the number of international migrants and refugees who had fled conflict, crises, persecution, violence or human-rights violations doubled from 17 million to 34 million. (This was all, of course, before Russia’s invasion of Ukraine forced more than four million refugees to flee the country.) Most of those compelled to move by the end of 2020 were internally displaced, as Sciubba notes. Between 2008 and 2020, an average of 21.8 million people per year had to move within their countries because of weather-related disasters, such as floods or wildfires.

Without future immigration, Sciubba writes, Europe’s population is set to shrink from about 521 million to about 482 million by mid-century. Immigrants could replace retiring workers. But while poorer countries including Jordan, Turkey and Kenya host 85% of forcibly displaced people, “richer countries go to great lengths” to keep out refugees and asylum seekers from war-torn countries including Syria, South Sudan and Afghanistan, Sciubba writes. She devotes little space to exploring the causes of this xenophobia, or how both conventional media and social media

frequently frame immigrant people as a threat. She merely notes that change — “particularly demographic change” — is hard on a society.

Even as populations rise in countries in sub-Saharan Africa, it’s worth noting that their consumption of resources is often minuscule compared with that in countries with ageing populations. According to a report by the Center for Global Development in Washington DC, by the end of the first week of January 2022, the average person in the United States had surpassed the annual carbon dioxide emissions of equivalent citizens in 22 low-income countries (see [go.nature.com/3lbhqju](https://go.nature.com/3lbhqju)). “In the long run, fewer wealthy consumers will likely be a net positive for Earth,” Sciubba writes.

“Unsustainable consumption and population growth are choking the seas with plastic, clogging our rivers with chemicals, and filling the planet’s atmosphere with greenhouse gases.”

And doom-laden prophecies do not always come to pass. In their 1968 book *The Population Bomb*, Paul and Anne Ehrlich argued that humans were breeding themselves to extinction, and warned of the perils of mass starvation. Thanks to humanitarian intervention, global health campaigns, vaccination and improvements in sanitation, deaths from communicable diseases and famine fell abruptly in the decades before the COVID-19 pandemic. Deaths from HIV declined by 51% from 2000 to 2019, with the use of antiretroviral therapy saving millions of lives. Starvation in countries such as Yemen (or Afghanistan in 2022, where more than half the population is “acutely food insecure”, according to the UN World Food Programme) is not solely the result of population increase. It is often the outcome of economic sanctions and, in the case of Yemen, the bombing of infrastructure including markets, roads, ports, farms and water pipelines.

“A good health foundation makes countries far more prepared for the unexpected,” Sciubba concludes. With the right political will and with humanitarian intervention, it seems, people can turn the tide on disease and war.

*Nature* **604**, 33-34 (2022)

doi: <https://doi.org/10.1038/d41586-022-00926-6>

## Updates & Corrections

- **Correction 07 April 2022:** An earlier version of this book review erroneously stated that one-quarter of the Japanese population was expected to have dementia by 2045. In fact, the proportion refers only to people aged older than 65.

---

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-022-00926-6>

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

- BOOK REVIEW
- 11 March 2022

# Taming stars, and how do birds smell? Books in brief

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

- [Andrew Robinson](#) 0



**ITER and the International Quest  
for Fusion Energy**

# **STAR POWER**

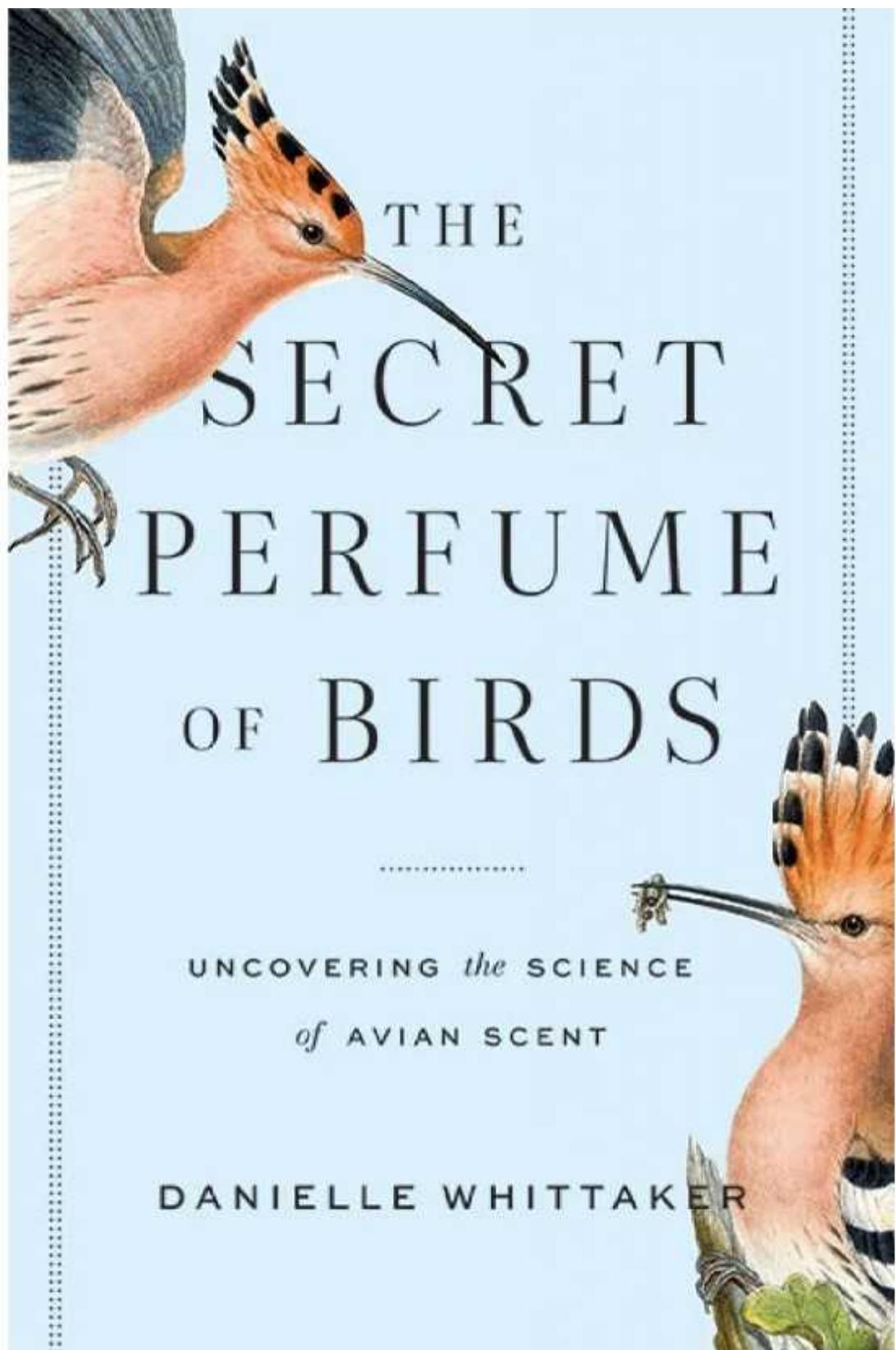
**Alain Bécoulet**

translated by Erik Butler  
foreword by Dennis G. Whyte

**Star Power**

Alain Bécoulet (transl. Erik Butler) *MIT Press* (2022)

The experimental nuclear fusion reactor ITER is probably the most expensive scientific experiment ever. Construction began in France in 2013; completion is scheduled for 2025. Alain Bécoulet, ITER's head of engineering, captures some 60 years of research in "the somewhat mad enterprise of domesticating the energy of the stars". Fusing one gram of hydrogen into helium makes as much energy as burning eight tonnes of petroleum. Despite many technical challenges, results convince Bécoulet that the star-makers will succeed.



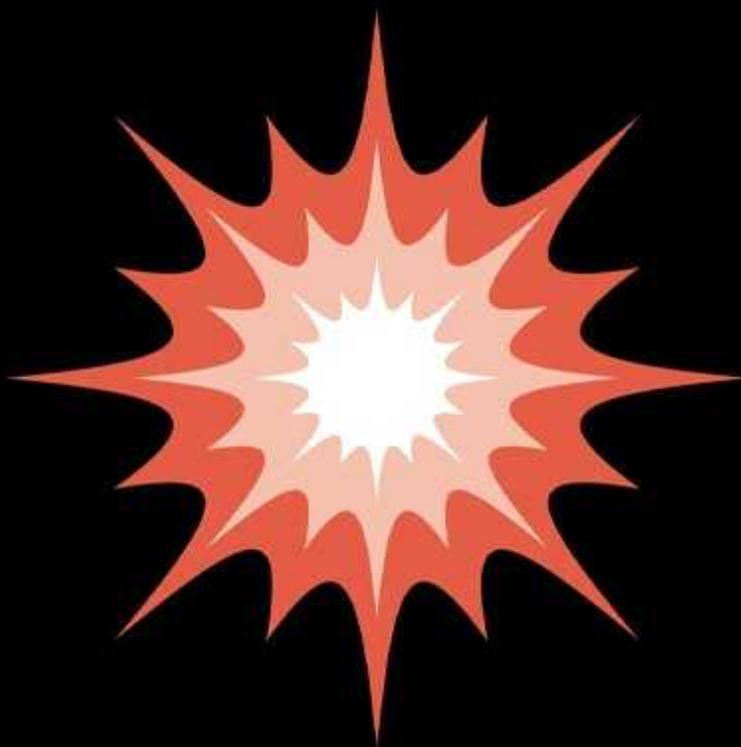
The Secret Perfume of Birds

Danielle J. Whittaker *Johns Hopkins Univ. Press* (2022)

A neurobiologist once told Danielle Whittaker that birds have no sense of smell, as if it were common knowledge, unworthy of her postdoctoral research. Unpersuaded, she pursued the subject. She found the theory's origin in ornithologist John Audubon's nineteenth-century experiments with vultures. Her study — mixing memoir, science and intellectual history — is a charming journey into “not just beaks, bulbs and behaviors” but “immune systems, hormones, genes, and bacteria”. Smelly chemicals made by birds prove vital to their lives.

# SUPERNova

OR GRAUR



THE MIT PRESS ESSENTIAL KNOWLEDGE SERIES

**Supernova**

*Or Graur MIT Press (2022)*

Supernovae — exploding stars — were first observed as far back as ad 185, by Chinese astronomers. In 1998, their use to measure distances to faraway galaxies enabled the discovery that the Universe's expansion is accelerating — part of an astrophysics revolution. “I sincerely hope that in ten years, twenty at most, this book will be woefully out of date,” writes astrophysicist Or Graur in his accessible introduction. It includes a periodic table coloured to show chemical elements formed from supernovae, such as calcium and iron.

# Otherlands



A World in  
the Making

Thomas  
Halliday

Otherlands

Thomas Halliday *Allen Lane* (2022)

Palaeontologist and biologist Thomas Halliday's history of complex life begins near the end, in the Pleistocene epoch 20,000 years ago, and travels back to the Ediacaran period 550 million years ago. Each chapter evokes a single site, ranging from Alaskan plains once dominated by mammoths to Australian hills that carry the traces of microbial mats. This imaginative — if stingily illustrated — portrayal of “a series of worlds, simultaneously fabulous yet familiar” is ever fossiliferous but never fossilized.

# The Birth of Modern Neuroscience in Turin

Stefano Sandrone & Lorenzo Lorusso



**The Birth of Modern Neuroscience in Turin**

Stefano Sandrone & Lorenzo Lorusso (eds) *Oxford Univ. Press* (2022)

The city of Turin in Italy has seen key contributions to neuroscience, starting with early-nineteenth-century anatomical drawings by Luigi Rolando. During the Second World War, Nobel laureate Rita Levi-Montalcini set the stage for her discovery of nerve growth factor in her bedroom laboratory, after being dismissed from Turin University under anti-Jewish laws. They are among many figures in this pioneering study edited by neuroscientists Stefano Sandrone and Lorenzo Lorusso — neither of whom works in Turin.

*Nature* **604**, 34 (2022)

*doi:* <https://doi.org/10.1038/d41586-022-00716-0>

---

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-022-00716-0>

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

# Opinion

- **[Crypto and digital currencies — nine research priorities](#)** [ 04 April 2022]  
Comment • To avert privacy breaches, scams and environmental damage, governments and central banks need to know how best to regulate this financial frontier.
- **[Russian scientists' complicity: shame will remain](#)** [ 05 April 2022]  
Correspondence •
- **[Regreening: green is not always gold](#)** [ 05 April 2022]  
Correspondence •
- **[Kharkiv's physicists, under fire, call for support](#)** [ 05 April 2022]  
Correspondence •
- **[China: protect black soil for biodiversity](#)** [ 05 April 2022]  
Correspondence •

- COMMENT
- 04 April 2022

# Crypto and digital currencies — nine research priorities

To avert privacy breaches, scams and environmental damage, governments and central banks need to know how best to regulate this financial frontier.

- [Andrew Urquhart](#) <sup>0</sup> &
- [Brian Lucey](#) <sup>1</sup>



Bitcoin is now a legal currency in El Salvador. Credit: Jose Cabezas/Reuters

Money is at a crossroads. A race is on to decide who creates it, who can access it and how, who controls it, and to what degree and how it is regulated. The outcome could decide whether governments have access to all our financial data, whether criminals can easily launder vast sums unseen, and whether the benefits of finance can be extended to the billions of people globally who lack access to banks.

Cryptocurrencies — private monetary systems that use digital encryption to perform and verify transactions — have been around since Bitcoin was developed in 2008<sup>1</sup>. Now, nearly 20,000 cryptocurrencies are in circulation, with a total market value of around US\$2 trillion.

Institutional investors are accumulating them. In the United States, many have dabbled publicly in this market, including: life-insurance company MassMutual; investment bankers JP MorganChase; hedge-fund manager Paul Tudor Jones; business-intelligence company MicroStrategy; and electric-vehicle manufacturer Tesla. More than 150 cryptocurrency hedge funds have sprung up to invest billions of dollars solely in these assets. The financial system is adapting, with bitcoin futures, options and exchange-traded funds now available. Coinbase, the leading exchange, is listed on the NASDAQ stock exchange.

Cryptocurrencies can potentially be used to hedge against inflation. Inflation fears are being stoked by the falling value of conventional currency. Governments and central banks deliberately expanded the money supply to invigorate economies during the 2008 global financial crisis and the COVID-19 pandemic. By contrast, the maximum supply of bitcoins is fixed at 21 million bitcoins.

Cryptos can offer financial independence. In 2021, El Salvador became the first nation to accept Bitcoin as a parallel official currency to the US dollar. The country's president, Nayib Bukele, thinks that cryptocurrencies can deepen financial inclusion and reduce economic dependence on the United States. Last year, the nation deposited US\$30-worth of bitcoins into each citizen's digital wallet, around one-tenth of the industrial minimum monthly wage. Technical issues plagued the launch, however, including hacked accounts, verification difficulties, poor information-technology

infrastructure and delays. In the past month, Ukraine has begun accepting donations in cryptocurrency to fund its defence against the Russian invasion.

Yet, in our view, great challenges need to be addressed before electronic currencies can become mainstream.

Bitcoin's value is notoriously volatile: 2021 saw swings of 50–100%. Vast amounts of energy are needed to run the complex computer codes that underpin cryptos. Currently, Bitcoin alone has an estimated annual carbon footprint equal to that of Kuwait and an electricity footprint equal to Thailand's, and produces as much electronic waste as the Netherlands. To get around this, El Salvador uses electricity generated by geothermal power near a volcano to decrypt its bitcoins. Lack of regulation also worries bankers and policymakers.

For all these reasons, China banned the mining of bitcoins last July. That August, the chair of the US Securities and Exchange Commission (SEC) warned that cryptocurrency platforms need regulation to survive. And in January this year, the UK Financial Conduct Authority (FCA) expressed concern over its inability to control cryptocurrency exchanges.

In parallel, several economies, including China and the European Union, are developing forms of digital assets over which they do have control. Central bank digital currencies (CBDCs) are an electronic record of the official currency, and might replace physical banknotes and coins within a decade. China is already testing a digital version of its yuan, with which customers can make payments over their mobile phones. Last July, the EU launched a project that will examine the feasibility of a digital euro over the next two years. Later this year, the US Federal Reserve will release a discussion paper on a digital dollar.

## Nine priorities

Although there's been much progress<sup>2</sup>, research is urgently needed on four fronts: legality, scalability, usability and acceptability. These are interconnected; solutions in one area might exacerbate problems in another. Here we set out nine priorities.

**Curb criminality.** The anonymity of cryptocurrency transactions means that they appeal to criminals and bad actors. Cryptos are used to launder money, fund terrorism and fuel corruption<sup>3</sup> — it's been estimated that up to half of bitcoin transactions could support illegal activities<sup>4</sup>. Cryptocurrencies might be used to bypass financial sanctions, such as those currently imposed on Russia.

Solving this problem is close to impossible — cash has been used to carry out all the crimes for which cryptocurrency is blamed. But options could be considered. Cryptocurrencies could monitor users and introduce ‘know-your-customer’ forms of authentication, a feature absent up to now. But that would go against their principles of privacy and ownership, lowering their uptake. Researchers need to examine such trade-offs.

CBDCs have the opposite problem. They could allow central banks and governments to track everyone’s transactions — a digital panopticon. However, if implications for privacy and rights were addressed, such transparency might deter and expose crime, as well as increasing tax revenues and diminishing black-market trade.

**Regulate digital assets.** At one extreme, cryptocurrencies are banned, as in China, Iraq and Egypt. At the other, they are mainstream, as in El Salvador. In most other countries they are permitted, yet not seen as legal tender. Trading is unregulated and they are treated as just another asset. Scams and misselling are rife.

Researchers need to ask: what is the optimal regulatory framework for digital assets to prosper in? And, how can regulators work in this space to the benefit of all?

Major global regulators such as the SEC in the United States and the FCA in the United Kingdom would like to see cryptos regulated much like other financial assets. Only qualified investors and authorized brokers should trade. Again, that would make cryptos less attractive, potentially crashing the market. Private currencies have existed before. In medieval times, and in nineteenth-century United States, Sweden and Australia, for example, individuals and banks could issue currency. These were all regulated out of existence.

## Key terms

A money glossary.

**Blockchain.** A distributed database that registers information electronically in a digital form. Data are structured into blocks that are linked to previous ones through cryptographic codes.

**Cryptocurrencies.** Digital assets secured by cryptography. Many, including Bitcoin, use blockchain technology to store transactions.

**Central bank digital currencies.** The virtual format of a fiat (official) currency: an electronic record, governed and regulated by state or federal monetary authorities.

**Fiat currency.** A national currency that is not pegged to the price of a commodity such as gold or silver. Central banks or governments control how much is issued. The US dollar is one example.

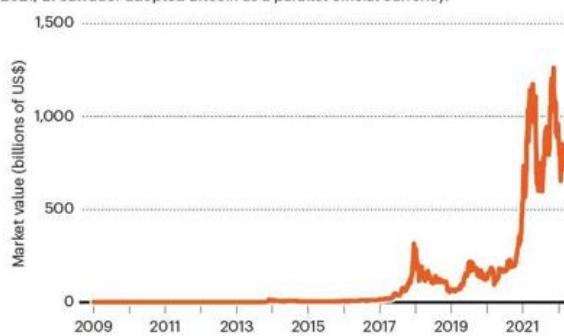
**Cut energy use.** Cryptocurrencies are computationally intensive (see ‘Bitcoin balance sheet’). Creating (‘mining’) some cryptocurrencies requires solving difficult cryptographic puzzles. Adding transactions to a digital ledger, such as the blockchain (see ‘Key terms’), demands verifications by algorithms. All those calculations consume energy. A report from the UK financial site MoneySuperMarket.com found that bitcoin is more energy hungry than Norway. The report found that processing one transaction consumes more than \$100 worth of electricity, and generates more than 800 kilograms of carbon dioxide (see [go.nature.com/3wzhkmw](https://go.nature.com/3wzhkmw)). The next most popular cryptocurrency, Ethereum, consumes less energy but still adds 62 kilograms of CO<sub>2</sub> per transaction.

## BITCOIN BALANCE SHEET

Since 2008, the cryptocurrency Bitcoin has blossomed into a virtual global currency. But steps still need to be taken to avoid price spikes and crashes and to lessen the computing power required.

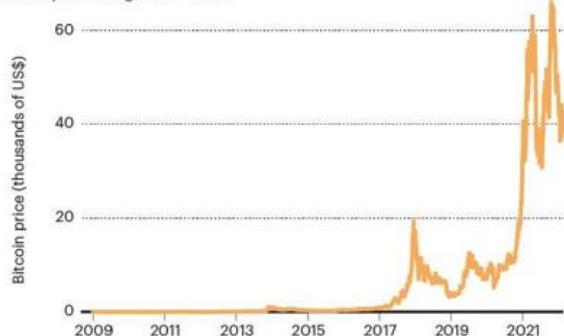
### Trillion-dollar market

In 2021, El Salvador adopted Bitcoin as a parallel official currency.



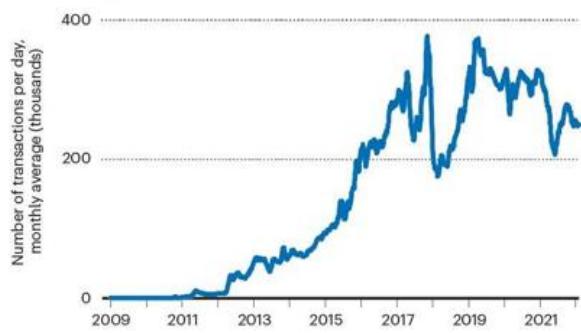
### Volatile prices

2021 saw price swings of 50–100%.



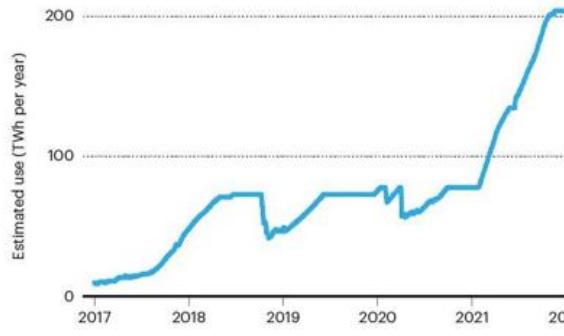
### Rising transactions

Bitcoin exchanges are slow and can take tens of minutes.



### High energy use

Bitcoin's carbon footprint has more than doubled in less than a year.



Sources: Energy, Digiconomist.net; All else, Blockchain.com

These high rates stem mainly from the choice of ‘consensus mechanism’ — all miners must agree on the computed solution to verify a transaction. Bitcoin uses a proof-of-work (PoW) system. If a miner is successful, they can propose a new block of transactions to the blockchain, and receive a reward. The reward diminishes in value but is still high: it started out at 50 bitcoins and halves every 210,000 blocks (roughly every 4 years), reaching 6.25 bitcoins for each block created in 2022. Lots of miners compete for the reward, hiking up energy use.

Shifting to an alternative consensus process — proof of stake (PoS) — could reduce energy consumption one-million-fold. Miners compete instead on the basis of their holdings of the coin. Ethereum is moving to PoS during 2022. But regulation and taxation risk disincentivizing PoS coins.

More research is needed on how incentives and sanctions can reduce energy needs. Options range from outlawing profligate protocols to nudges towards more efficient ones through regulatory or tax favouritism.

**Speed up transactions.** To replace existing systems of payment, cryptocurrencies will need to challenge retail and commercial banking settlement systems (such as SWIFT). These process payments in seconds and handle quadrillions of dollars per year. By contrast, Bitcoin handles 250,000 confirmed transactions per day, with a market value of \$1 billion to \$2 billion. Its transactions can take tens of minutes — much too slow for global market needs. Researchers need to find ways to speed them up.

A handful of cryptocurrencies can handle large volumes of transactions. EOS, for example, can manage 50,000 transactions per second. But it is a centralized cryptocurrency, less private than Bitcoin.

Volatility adds complications. In any monetary transaction, both sides want certainty in the value of an asset<sup>4</sup>. Yet bitcoin values can vary by 10% in minutes. Few will sell goods if they lose a big chunk of value between sale and settlement.

Bypassing some of the cumbersome blockchain technology goes part of the way to solving the speed problem. Adding a second layer to a network can enable transactions off the blockchain. Bitcoin, for example, introduced such a ‘lightning network’ in 2018, although it has only recently become popular. It allows two parties to make or receive payments swiftly off-chain, while the transactions are logged.



Employees work on bitcoin-mining computers in a Bitminer factory in Italy. Credit: Alessandro Bianchi/Reuters

**Manage volatility.** Cryptocurrency prices can have massive swings<sup>5</sup>. Last year, Bitcoin rose from \$37,000 in January to \$64,000 in April, falling back to \$37,000 by July then rising again to \$67,000 in November. That June, a mid-sized coin (Iron Titanium) collapsed from \$51 to \$0.0004 in 24 hours. There are many reasons for volatility, including these assets’ immaturity, decentralized trading and limited supply.

Mainstream currencies don’t fluctuate so much because central banks buy and sell currency to moderate swings. Other financial products such as options and futures also smooth price paths because these allow investors to hedge future risks. The introduction of these instruments for

cryptocurrencies has not reduced volatility, however<sup>6</sup>. Cryptocurrency exchanges lack limits and rules to halt trading when prices rise or fall more than a set percentage over a period; such rules act as an emergency brake to slow price crashes or bubbles.

Researchers should study factors driving cryptocurrency volatility and how to manage it. The relationship of cryptocurrencies to other assets in diverse investment portfolios needs attention, as does the impact of price volatility and policy uncertainty on Bitcoin<sup>7</sup>. Regulators should provide warnings about price swings to investors.

**Boost security.** Cash can be lost, credit cards stolen and bank fraud committed. Sometimes the holder is insured or compensated by an insurance scheme. But if the electronic keys to a cryptocurrency wallet are lost, there's no way to gain access and no central body to help users. Some estimates suggest that up to 25% of all bitcoins ever mined — hundreds of billions of dollars' worth — is irretrievable.

Virtual raids are common. For example, in 2014, the Japan-based bitcoin exchange Mt. Gox was hacked. At the time, it handled 70% of bitcoin trading; \$473 million dollars' worth of bitcoins were stolen.

Researchers need to explore how cryptocurrency can better withstand cyberattacks. State-run CBDCs should be more secure, but keeping them so will require effort. Exchanges should offer education around security on their platforms. More might require user identification.

**Manage fees.** Cryptocurrencies charge fees for transactions. The person making the transaction sets the amount and miners naturally want to work on those cryptocurrencies with the highest fees. Competition drives up these fees (varying from one cent to more than \$50) and fee spikes are common when many transactions are in the queue needing to be verified<sup>8</sup>. This typically happens when prices tumble and users try to offload coins, as happened with Bitcoin in early 2018.

Researchers need to examine how to control and smooth such spikes. They should set out guidelines on how users choose their fees. Miners should be

encouraged through regulation and incentives to add transactions to the next block on the basis of the timestamp and not the fee.

**Educate users.** A 2021 study by the FCA found that 78% of UK consumers have heard of cryptocurrencies. Few know how they work. Around 10% understood wrongly that they had consumer protection for their crypto holdings, as they might for bank deposits. More than 40% relied on information from social media; ownership was concentrated in younger people. Most finance experts lack training in the intricacies of ‘fintech’. Universities have only recently started offering modules on cryptocurrencies and blockchain.

Media coverage tends to be negative — focused on lost wallets, criminality, volatility and energy use. Few merchants accept cryptocurrencies, leaving them as new instruments for speculators<sup>9</sup>. Businesses and educators should do more to improve customers’ understanding and build trust.

**Protect privacy.** CBDCs come with a big privacy challenge, unlike decentralized cryptos. A move to a cashless society could allow governments to see all our transactions. Given growing interest from central banks, research is urgently needed on the willingness of individuals to trade privacy for convenience in their financial transactions<sup>10</sup>.

Baseline surveys would be a first step. China is already road testing its e-yuan. Since 2020, the People’s Bank of China has been randomly gifting customers e-yuan, which they can use to purchase goods and services. A prepaid card and e-banking app was trialled in February at the Beijing Winter Olympics.

More work on all these aspects will help governments and the public decide what the future of money should look like.

*Nature* **604**, 36-39 (2022)

*doi:* <https://doi.org/10.1038/d41586-022-00927-5>

## References

1. Nakamoto, S. *Decentralized Business Rev.* **21**, 260 (2008).
2. Corbet, S., Lucey, B., Urquhart, A. & Yarovaya, L. *Int. Rev. Finan. Anal.* **62**, 182–199 (2019).
3. Teichmann, F. M. J. & Falker, M.-C. *J. Money Laundering Control* **24**, 775–788 (2020).
4. Foley, S., Karlsen, J. R. & Putnins, T. J. *Rev. Finan. Stud.* **32**, 1798–1853 (2019).
5. Shen, D., Urquhart, A. & Wang, P. *Eur. Finan. Manage.* **26**, 1294–1323 (2020).
6. Corbet, S., Lucey, B., Peat, M. & Vigne, S. *Econ. Lett.* **172**, 23–27 (2018).
7. Lucey, B. M., Vigne, S. A., Yarovaya, L. & Wang, Y. *Finan. Res. Lett.* **45**, 102147 (2022).
8. Easley, D., O’Hara, M. & Basu, S. *J. Finan. Econ.* **134**, 91–109 (2019).
9. Baur, D. G., Hong, K. & Lee, A. D. *J. Int. Finan. Markets Inst. Money* **54**, 177–189 (2018).
10. Agur, I., Ari, A. & Dell’Ariccia, G. *J. Monet. Econ.* **125**, 62–79 (2022).

---

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-022-00927-5>

- CORRESPONDENCE
- 05 April 2022

# Russian scientists' complicity: shame will remain

- [Leonid Gurvits](#) 0

Among the many victims of Russian President Vladimir Putin's barbaric invasion is science in Ukraine, Russia and elsewhere. As in the 1930s and 1940s, researchers in the aggressor country face a moral dilemma: to support the war and become pariahs among their global peers, or to oppose it and face persecution at home. Scientists bear special responsibility for making the right choice, having the tools and skills to distinguish between information and propaganda, and to stay fully informed.

## Access options

Subscribe to Nature+

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

\$29.99

monthly

[Subscribe](#)

Subscribe to Journal

Get full journal access for 1 year

\$199.00

only \$3.90 per issue

## Subscribe

All prices are NET prices.

VAT will be added later in the checkout.

Tax calculation will be finalised during checkout.

Buy article

Get time limited or full article access on ReadCube.

\$32.00

## Buy

All prices are NET prices.

## **Additional access options:**

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

*Nature* **604**, 40 (2022)

*doi:* <https://doi.org/10.1038/d41586-022-00941-7>

---

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-022-00941-7>

- CORRESPONDENCE
- 05 April 2022

# Regreening: green is not always gold

- [Michael C. Orr](#) <sup>0</sup> &
- [Alice C. Hughes](#) <sup>1</sup>

As the upcoming United Nations Biodiversity Conference in Kunming, China, ushers in the UN decade of ecosystem restoration, regreening efforts are sprouting worldwide. Adding vegetation — expedited by new technologies such as EcoFit, which predicts what trees will thrive in a given environment — can salvage highly disturbed habitats, benefiting native species and offsetting climate change. But when aimed at halting desertification, regreening can have a devastating cost for native ecosystems.

## Access options

Subscribe to Nature+

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

\$29.99

monthly

[Subscribe](#)

Subscribe to Journal

Get full journal access for 1 year

\$199.00

only \$3.90 per issue

[Subscribe](#)

All prices are NET prices.

VAT will be added later in the checkout.

Tax calculation will be finalised during checkout.

Buy article

Get time limited or full article access on ReadCube.

\$32.00

[Buy](#)

All prices are NET prices.

### **Additional access options:**

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

*Nature* **604**, 40 (2022)

*doi:* <https://doi.org/10.1038/d41586-022-00944-4>

---

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-022-00944-4>

- CORRESPONDENCE
- 05 April 2022

# **Kharkiv's physicists, under fire, call for support**

- [Yuri Naidyuk](#)<sup>0</sup>,
- [Olexander Dolbin](#)<sup>1</sup>,
- [Mykola Glushchuk](#)<sup>2</sup> &
- [Vlada Pashynska](#)<sup>3</sup>

Russia's barbarous invasion of Ukraine is, among many horrors, destroying our scientific infrastructure and community, and using disinformation to undermine global trust in research.

## **Access options**

Subscribe to Nature+

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

\$29.99

monthly

[Subscribe](#)

Subscribe to Journal

Get full journal access for 1 year

\$199.00

only \$3.90 per issue

## Subscribe

All prices are NET prices.

VAT will be added later in the checkout.

Tax calculation will be finalised during checkout.

Buy article

Get time limited or full article access on ReadCube.

\$32.00

## Buy

All prices are NET prices.

## **Additional access options:**

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

*Nature* **604**, 40 (2022)

*doi:* <https://doi.org/10.1038/d41586-022-00943-5>

---

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-022-00943-5>

- CORRESPONDENCE
- 05 April 2022

# China: protect black soil for biodiversity

- [Deyi Hou](#) ✉

In December 2021, the National People's Congress of China released a draft law on the protection of black soil, noted for its high humus and nutrient content and strong structure. To align with the post-2020 Global Biodiversity Framework under discussion at the United Nations Biodiversity Conference (COP-15) in Kunming, China, later this year, the soil law and the national action plan on black-soil protection must be strengthened to include specific and measurable requirements for biodiversity protection.

## Access options

Subscribe to Nature+

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

\$29.99

monthly

[Subscribe](#)

Subscribe to Journal

Get full journal access for 1 year

\$199.00

only \$3.90 per issue

[Subscribe](#)

All prices are NET prices.

VAT will be added later in the checkout.

Tax calculation will be finalised during checkout.

Buy article

Get time limited or full article access on ReadCube.

\$32.00

[Buy](#)

All prices are NET prices.

### **Additional access options:**

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

*Nature* **604**, 40 (2022)

*doi:* <https://doi.org/10.1038/d41586-022-00942-6>

---

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-022-00942-6>

# Work

- **Time to rethink the scientific CV** [ 05 April 2022]  
Career Feature • Fresh formats showcase researchers' work more effectively.
- **In pursuit of data immortality** [ 04 April 2022]  
Technology Feature • Data sharing can save important scientific work from extinction, but only if researchers take care to ensure that resources are easy to find and reuse.
- **Net benefit: using a turtle excluder device in the Adriatic Sea** [ 04 April 2022]  
Where I Work • Marine biologist Laura Aiudi is working on a net that saves the lives of endangered species — but still protects the livelihood of fishermen.

- CAREER FEATURE
- 05 April 2022

# Time to rethink the scientific CV

Fresh formats showcase researchers' work more effectively.

- [Chris Woolston](#) ✉

[Find a new job](#)



New CVs formats allow researchers to highlight contributions beyond their publication list.Credit: Getty

In December 2021, UK Research and Innovation (UKRI), the largest public funder of UK science, announced that it was abandoning the use of the conventional CV — curriculum vitae — in funding applications. The

funding body said it would adopt a new type of CV to “enable people to better demonstrate their contributions to research, teams, and wider society”.

As institutions and funders around the world [reassess their approach to researcher evaluations](#), there’s a growing call to revamp the academic CVs used to support applications for jobs, funding, promotions and awards. Researchers need to find fresh ways to document their accomplishments and value beyond a mere listing of publications, and committees overseeing promotions and grants need to change their protocols and expectations, says Needhi Bhalla, a cell biologist at the University of California, Santa Cruz. “CVs should reflect the authentic experience of being a scientist,” she says, including mentorship, work on committees, outreach and many other contributions that don’t result in publications. “I’m excited that we’re in the process of rethinking them.”

CVs have long been part of the currency of scientific promotion. Scientists seeking a position or a grant often feel obliged to list every publication, presentation and award in a single document intended to sway committees through its sheer length and volume. The typical CV follows a time-worn template, says Robert Morrell, an education researcher and former director of the New Generation of Academics Programme at the University of Cape Town in South Africa. “ ‘I was born, I went to school here, I had these publications, these are the students I graduated.’ People who write CVs like that are missing the boat.”

The UKRI is not alone in seeking to rethink the CV in response to a renewed focus on team science and equity, diversity and inclusion (EDI). It modelled its new CV format on [‘Résumé for Researchers’](#), introduced in 2019 by the Royal Society in London. Similar initiatives have been unveiled by research councils in the Netherlands and Luxembourg.

In response, researchers are learning how to rework CVs to emphasize quality over quantity, and to include narratives about their broader impact. Meanwhile, hiring panels and grant evaluators need to rethink how best to assess these documents.

The core problem with standard CVs is that they tend to reduce scientists to numbers, says Rebecca Pillai Riddell, a behavioural scientist and associate

vice-president of research at York University in Toronto, Canada. Evaluating researchers on the basis of sheer number of publications or using related measures, such as the impact factors of the journals in which they publish, ignores many things that go into a scientific career, Pillai Riddell says. Conventional CVs “are supposed to be quick-and-dirty summaries”, she says. As someone who has seen many over the years, she knows that those summaries can contain valuable information, even if the emphasis is often misplaced. “They focus on counting, not on what’s important.”

The ‘quantity above quality’ approach is especially short-sighted and unfair in the wake of the COVID-19 pandemic, Pillai Riddell says. Many researchers simply didn’t have the time or opportunity to conduct experiments or crank out papers at their normal pace during shutdowns. And as schools closed their doors, many scientists who were also parents had to shift their priorities from work to home, especially women. “If we continue to emphasize quantity, caregivers are not going to be eligible for grants or awards,” she adds.

Scientists and institutions alike need to reconsider the entire purpose of a CV, says Wolfgang Kaltenbrunner, a sociologist of science at Leiden University in the Netherlands. “To make science work, you need to accomplish a lot of tasks that are not easily represented in a CV,” he says, such as communicating science to the general public and collaborating behind the scenes on big projects. “Are we selecting for the right things in grant funding or tenure? There’s widespread discontent with it in science.”



Robert Morrell (in purple top) on a retreat with early-career researchers in South Africa.Credit: Robert Morrell

## Contributions that count

Kaltenbrunner co-authored a 2021 commentary in *Humanities & Social Science Communications* that offered ten suggestions for revamping academic CVs to make them a fairer gauge of scientific talent<sup>1</sup>. They include a new focus on “activities and outputs that are relevant”. That means moving away from exhaustive lists of publications and presentations, and cutting down on ‘noise’ that doesn’t reflect qualifications for a job or grant. Instead of including everything that has ever carried their name, researchers should list a few meaningful publications that hiring managers and evaluators could realistically take the time to read and appreciate, Kaltenbrunner and his colleagues say. “Focusing on only a few outputs saves researcher and evaluator resources, discourages salami slicing of results, improves comparison between early- and late-career researchers and renders

publication hiatuses as a result of career breaks less apparent,” they write. Importantly, such an approach would help to level the playing field when early-career and senior scholars are directly competing.

Pillai Riddell would welcome a résumé revolution that cuts down on reading for those who assess applicants. “I’m thinking about reviewer burdens,” she says. “In my dream scenario, you’d pick two papers and provide a 200-word summary of the importance of the paper. It allows for contextualization.”

Kaltenbrunner notes that many academic jobs require a covering letter, which gives applicants another opportunity to tell the story of their careers and highlight their most important papers. “They can use the narrative to fill gaps that are left by the publication record,” he says.

Publication lists aren’t as meaningful today as they might have been for previous generations of scientists, Kaltenbrunner says. “Science has become increasingly competitive in the past 40 years, so the publication lists have become much longer,” he says. “It’s not necessarily true that people have more ideas, but publication conventions have changed. Competition actually reduced the informational value of CVs.”

Appraising someone according to their number of publications and how many times these have been cited also greatly favours researchers in particular fields, Pillai Riddell says. “If you’re studying bird mating calls instead of cancer, you aren’t going to have the same number of hits,” she says. She adds that medical researchers can show up on 20–30 papers a year, an impossible standard for someone in a field such as behavioural science. Such comparisons can become important in the context of international awards that attract applicants from across the scientific spectrum. Grant-awarding bodies should embrace diversity of scientific fields as well as other forms of diversity, she says.

CVs could be more effective if they allowed room for narratives — brief statements that tell a story about a scientist, their accomplishments or their impact. “A narrative section would give them room to explain their achievements and contributions to science that do not fit traditional CV categories,” Kaltenbrunner says. With a narrative section, “they could tell stories of successful engagement with a stakeholder, contributions in terms

of community service, or excellence in teaching or supervision”. (See ‘CV snapshots’ for examples.)

## CV snapshots

These excerpts are from the narrative CVs of successful applicants to the Luxembourg National Research Fund in 2021.

- “Alongside scientific goals, I also follow leadership ones. A four-day professional leadership course and three months of personal coaching in 2020 taught me to reflect on myself, develop my scientific vision and learn about key attributes of successful teams. I also sent my postdocs on similar courses. As a result, my team is extremely productive, with two manuscripts at the submission stage only 2.5 years after the launch of my own group.”
- “I give regular talks at foundations, charity clubs and student associations, telling young people about scientific research and new therapeutic avenues in cancer. I also regularly write for national newspapers, again to transfer my passion for research to younger people.”
- “I invest in the development of individuals and build up a strong team spirit by regularly taking leadership and conflict-management courses. During the COVID-19 pandemic, I put into practice various ideas on remote leadership and team communication.”
- “I made a 52-minute documentary about contemporary psychiatry in my country, together with a visual anthropologist and a local production company. We worked as care assistants on a ward for three months before introducing a camera. The film proved to be a stimulating exercise in public engagement.”

The term ‘narrative CV’ is gaining traction, but Kaltenbrunner says he’s not actually a fan of that label. “It’s binary,” he says. “It suggests that a CV is either narrative or not narrative.” He prefers ‘contextual CV’: “It’s more about supplementing traditional CVs with other elements.” He notes that the use of alternative CV formats by research councils in the Netherlands and Luxembourg has dismayed some, more senior, researchers. “Some see these

experimental CV formats as an undue intervention by funders,” Kaltenbrunner says. “They have made a career based on existing criteria.”

The Luxembourg National Research Fund says [the narrative CV model](#), introduced last year, will “allow an applicant to be more fairly evaluated on their scientific vision, appropriate experience, and contributions to science and society”. Similarly, the Royal Society’s Résumé for Researchers is a narrative-based document that is focused on four key questions: how have you contributed to the generation of knowledge? How have you contributed to the development of individuals? How have you contributed to the wider research community? And how have you contributed to broader society?

## Documenting diverse work

A new era of CVs could help to promote diversity in science, Bhalla says. “Traditional metrics of what you’ve published, where you’ve published and who you’ve published with, are definite barriers to diversity, equity and inclusion,” she says. Scientists who might not have wowed evaluators in the past with their publications and impact factors would have a chance to explain their mentorship, outreach and committee duties — areas where women and people from minority ethnic groups and other under-represented demographics often excel (see ‘Building a standout CV’).

## Building a standout CV

When he was director of the New Generation of Academics Programme, a South African initiative to recruit a diverse cohort of promising scholars to academia, Robert Morrell worked to give young researchers an edge. In many cases, that meant helping them to build a CV that truly captured their skills and potential, says Morrell, who retired from the position in 2021. “My job was to help people get promoted.”

He encourages researchers to “harvest evidence” of their work and its impact. It’s especially important to keep track of things that can’t be measured easily, including positive feedback from students or collaborators.

“I urge people to keep [complimentary] e-mails and file them in a separate folder,” he says, such as messages of praise for participating in a big team project. “Those types of examples are really helpful, and people don’t think of [including] them. They think it’s immodest.”

Likewise, Rebecca Pillai Riddell, associate vice-president of research at York University in Toronto, Canada, always keeps meticulous track of time and effort spent on mentoring, teaching and serving on committees.

“Nobody is going to track it for me,” she says. “To survive in academia, to get leadership roles, you have to advocate for yourself.”

Pillai Riddell says it’s easier to build a case for yourself if you organize your work systematically. “You need to create structures,” she says. For example, as an advocate for under-represented students, she has set aside 2 hours of office time every week specifically for them. She doesn’t keep track of everything that’s discussed, but she can report how many students have dropped in. “It’s about getting credit for what you’re already doing.”

Equity statements have become an increasingly common requirement for CVs, and universities that require such statements have experienced greater diversity in subsequent recruitment<sup>2</sup>. Riddell recommends including such a statement even if it’s not required. “You can say, ‘My commitment to equity requires that I tell you about this.’”

In a 2019 article in *Molecular Biology of the Cell*, Bhalla laid out a series of strategies to improve equity in faculty hiring<sup>2</sup>. Among other things, she suggested augmenting conventional CVs with short statements that summarize an applicant’s research contributions during their graduate studies and as postdocs.

Bhalla says that there’s been some pushback in the science community against any sort of narrative sections on CVs. Some say that asking scientists to explain the impact or importance of their work provides an unfair advantage for people with strong communication skills while hindering people who might not be as persuasive, including scientists who speak English as a second or third language. But Bhalla says being able to explain one’s research is a fundamental part of being a scientist. “Those are skill sets

that you're going to need anyway if you're writing a grant," she says. "So that's one of the skills that we should be assessing."

Such messages, no matter how well crafted, will only work if evaluators are ready to accept them, Riddell says. She's encouraged by the growing number of institutions that have signed the San Francisco Declaration on Research Assessment, a framework that, among other things, [discourages the use of impact factors](#) in hiring and funding decisions.

In March, Pillai Riddell and her team at York University launched [POLARIS](#), an online training course to help members of the university's hiring and funding committees to update how they evaluate researchers. One of the course's main goals is to encourage participants to look beyond CVs to consider EDI issues in their decisions. [It includes modules](#) in which participants rank hypothetical candidates and discuss their choices with the EDI programme manager. The training also includes videos of experienced evaluators discussing best practice. Pillai Riddell says that some of the videos captured real-life arguments, signifying the tension and passion that goes into researcher evaluation. After completing the course, participants receive a certificate that, naturally, can be included on their CV. (A version of the course is available to external researchers as well.)

The current use of CVs also hinders the career progression of scientists in developing countries who must get by with limited resources and infrastructure, says Olumuyiwa Asaolu, an engineer at the University of Lagos in Nigeria. In a 2020 [opinion piece for the academic news site The Conversation](#), Asaolu called for a fresh approach to evaluating African researchers, including rethinking the CV. The emphasis on publications and impact factors is especially problematic, he says, partly because of costly publication fees. "It's not easy for Africans to publish in the big journals."

Asaolu, who completed a postdoctoral position at the University of Tennessee at Knoxville in the early 2000s, says that publishing while working in different countries has given him insight into disparities. "The response you get if your address is in Africa is not the same as the response or treatment you get if you're sending your manuscript from a Western institution."

In 2018, a web-based survey of 267 African researchers conducted by Asaolu and his colleagues underscored doubts about standard metrics<sup>3</sup>. Although the majority (59%) of respondents agreed that impact factor is a true measure of a journal's quality, only 40% agreed that publishing in journals with high impact factors should remain a major component of winning grants and promotions.

An approach to CVs that focuses more on real-world contributions — including projects that help local communities — and less on impact factors could help to level the playing field for African researchers who are applying for positions or grants overseas, Asaolu says. But he adds that young researchers can't take it on themselves to reinvent the system. As a mentor, he often assists others with their applications, and encourages them to follow existing instructions and templates as closely as possible. "Change has to be incremental," he says.

Incremental or not, changes to the format of CVs are inevitable, Pillai Riddell says. Early-career researchers can do their part by expanding their own definitions of what's worth listing and, more fundamentally, what it means to be a successful scientist. They can certainly mention a paper of theirs with thousands of downloads, but shouldn't ignore the impact they've had on their communities: "Both have a place."

*Nature* **604**, 203-205 (2022)

doi: <https://doi.org/10.1038/d41586-022-00928-4>

## References

1. Strinzel, M., Brown, J., Kaltenbrunner, W., de Rijcke, S. & Hill, M. *Humanit. Soc. Sci. Commun.* **8**, 251 (2021).
2. Bhalla, N. *Mol. Biol. Cell* **30**, 2744–2749 (2019).
3. Atolani, O. *et al.* in *Globafricalisation and Sustainable Development* <https://doi.org/10.2478/9783110671049-010> (Sciendo, 2019).

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-022-00928-4>

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

- TECHNOLOGY FEATURE
- 04 April 2022

# In pursuit of data immortality

Data sharing can save important scientific work from extinction, but only if researchers take care to ensure that resources are easy to find and reuse.

- [Michael Eisenstein](#) <sup>0</sup>

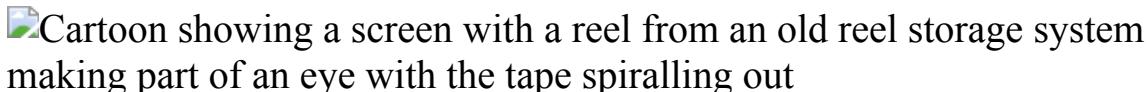


Illustration by The Project Twins

Between 1969 and 1972, the United States landed six crewed spacecraft on the Moon as part of the Apollo programme. The missions retrieved priceless samples. But for more than four decades, the data from those samples remained stashed away at a handful of US laboratories — until Kerstin Lehnert came along.

A geoinformatician specializing in data rescue and preservation, Lehnert set out in 2014 to transform these data sets into a usable resource. Her team at Columbia University's Lamont–Doherty Earth Observatory in Palisades, New York, pored through old conference abstracts, scanned reams of publications and debriefed the senior researchers who first studied those lunar samples to collect, organize and annotate as much information as possible. One scientist, Lehnert says, “came with a half-metre-high pile of old, folded printouts and we spent a whole summer typing those data into Excel spreadsheets”. Thanks to their efforts, these one-of-a-kind data are now freely available in the Astromaterials Data System.

Countless other laboratories, and their precious, irreplaceable data, are not so fortunate.

## Lost to the ages

‘Big science’ efforts led by international consortia typically have data-management and sharing plans built in. But many labs doing small- to medium-scale studies in more specialized areas — such as analysing the biological contents of a single lake, or tracking the physiology of specific animal models — have no such systems. Their data often remain siloed in the labs that generated them, fading from memory as project members leave.

For the scientific community, that’s a tragedy of wasted effort, lost collaborative opportunities and irreproducibility. “Things don’t have to be really popular in order to be still very valuable,” says Erik Schultes, international science coordinator for the GO FAIR International Support and Coordination Office in Leiden, the Netherlands. Established in 2018 to develop best practices for data preservation and sharing, GO FAIR is one of several efforts engaging with researchers in almost every scientific discipline to secure today’s data for posterity. But success will require a concerted effort — and a shift in lab culture.

Digital data might be more convenient and shareable than the paper notebooks and printed photographs of yore, but they won’t last forever. Physical storage media degrade; file formats and the software that produced them become obsolete. Most importantly, scientists can lose track of data when they stop being immediately useful. Even if retrieved, archival files often lack the context needed to interpret them.

“I’ve gone back and tried to make sense of data that I collected 10 or 15 years ago,” says Dominique Roche, an ecologist at Carleton University in Ottawa who also studies data reuse and reproducibility. “I’m particularly knowledgeable about proper data management, and it was almost impossible.” The difficulty only grows when researchers seek older data from other groups. In 2013, Timothy Vines, a data scientist then at the University of British Columbia in Vancouver, Canada, and his colleagues tested the limits of this accessibility by requesting data from 516 studies published between 2 and 22 years earlier. They managed to retrieve fewer

than one in 5 data sets, and found that the likelihood of data being available and usable dropped by 17% each year after publication<sup>1</sup>.

In recent years, researchers have taken to uploading their data to open-access repositories. This is an important step towards preservation and access, but it doesn't ensure reusability. In a survey of 100 data sets on the repository Dryad, Roche and his colleagues found that more than half lacked data needed to reproduce the work, and more than one-third were either not machine-readable or essentially unusable in other ways<sup>2</sup>.

This is assuming that one can even find a particular data set: shared data can be scattered among multiple repositories, and it can be challenging to search across them, says Schultes.

## A FAIR solution

The good news is that more sophisticated solutions are emerging. In 2016, a multinational team coordinated by Barend Mons, a specialist in biosemantics at Leiden University Medical Center, and including Schultes, published a framework known as the FAIR Data Principles<sup>3</sup>. The acronymic title describes its objective: that scientific data should be findable, accessible, interoperable and reusable.

Many of the framework's goals can be met through careful data curation and metadata creation. Metadata consist of documentation that describes a data set in a format that is both human- and machine-readable. They might, for example, describe the cell types and imaging parameters used in a microscopy experiment. That's essential information for third-party analyses, but also for finding the data. Other tools that can aid findability include re3data, developed by data-preservation organization DataCite, based in Hanover, Germany, which can help users to quickly narrow down which repositories are most likely to contain data relevant to their research. Google also offers a Dataset Search service, which can search across thousands of repositories to uncover specific data sets.

Metadata generation can create considerable work, but there are resources to expedite it. The Center for Expanded Data Annotation and Retrieval

(CEDAR) at Stanford University in California runs a platform that generates simplified forms to produce FAIR-compliant metadata. These can be uploaded to repositories alongside the data they describe. GO FAIR also regularly runs Metadata for Machines workshops, at which data specialists and domain-specific experts help researchers to generate well-crafted metadata.

## Fleshing out the record

Other efforts aim to preserve historic data sets. For example, Canada's nationwide Living Data Project trains and supports junior scientists to work with labs that have precious archival data from ecology or environmental science but lack the skills or resources to preserve them adequately. Roche, one of the project's coordinators, says the goal is to "organize the data, manage them properly and create the metadata so that then the data can be made public and are going to be understandable and reusable". The group has taken on more than 40 projects since 2020, salvaging one-of-a-kind research material, including 20 years of records of flora from Canada's Yukon tundra, and observations of bird populations from Tanzania's Serengeti region dating back to 1929.

But however old the data, preservation isn't a one-time task: to remain usable, raw scientific data must be maintained in formats that are compatible with contemporary hardware, software and operating systems. "You have to continue migrating data forward," says Christine Borgman, an information scientist at the University of California, Los Angeles. "As each new technology comes along, you've got to keep on upgrading every time."

That's a burdensome process, acknowledges Klaus Rechert, a computer scientist at the University of Freiburg in Germany. "For every data format, you need a migration tool," he says, "and the number of data formats is exploding." As an alternative, Rechert's team focuses on emulation — using software to replicate the hardware and operating system required to run old programs. This means that researchers can interact with old data sets using the original software. It has the added benefit of preserving the software itself, which is an important component of the scientific record.

But emulation can be technically challenging. So Rechert and his colleagues at the University of Freiburg have developed the Emulation-as-a-Service Infrastructure (EaaSI) — a cloud-based system that researchers can use to boot up antiquated systems. For example, a user who needs to run software originally designed for an old Apple or PC — or even older systems such as those produced by Commodore — can replicate that computing environment on any modern machine running Linux. The emulator's complexity is hidden behind a user-friendly interface, with technical components managed by the EaaSI team. “We currently do everything to automate it,” says Rechert. “We are able to analyse the data set and try to figure out what is the most appropriate software environment.”

## A culture of preservation

With better tools available, the trick now is to give researchers incentives to put in the extra effort — a task that entails overcoming long-entrenched views on how scientific effort is credited and rewarded. This is especially true in academia, where publications remain the coin of the realm. Even with the advent of services such as DataCite, which provide ways to cite data sets, funders and hiring committees tend to gloss over those contributions in a scientist’s CV. “Institutions don’t really care whether your data sets get cited,” says Roche.

Some major funders — including the US National Institutes of Health and Wellcome in London — have formal requirements for data management and sharing, and a number of journals make repository use a precondition. This can be a big incentive: Lehnert notes that when several major geoscience journals adopted the FAIR principles in 2019, submissions to the EarthChem Library data repository tripled. But there is little close oversight, and few teeth for punishing non-compliance; and researchers are rarely given the resources to support preservation efforts. “It keeps getting pushed down to the principal investigator as their responsibility,” says Borgman.

Remedying this will require structural changes in the infrastructure for scientific funding and support. But the rising generation of scientists —

born into an era of open-access, open-source and automated science — might be more amenable to the effort than their predecessors. “Nobody wants to hear that they might die tomorrow, but maybe your computer dies tomorrow and you don’t have a good back-up,” says Lehnert. “The data has to go into the repository so that 20 years from now, we’re not suddenly saying, ‘We need to invest again in rescuing these data.’”

*Nature* **604**, 207-208 (2022)

*doi:* <https://doi.org/10.1038/d41586-022-00929-3>

## References

1. Vines, T. H. *et al.* *Curr. Biol.* **24**, 94–97 (2014).
2. Roche, D. G., Kruuk, L. E. B., Lanfear, R. & Binning, S. A. *PLoS Biol.* **13**, e1002295 (2015).
3. Wilkinson, M. *et al.* *Sci. Data* **3**, 160018 (2016).

---

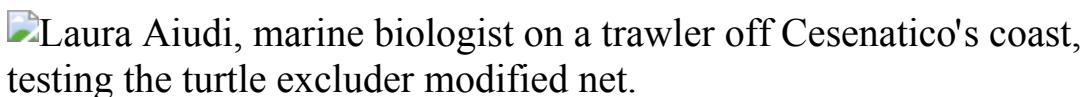
This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-022-00929-3>

- WHERE I WORK
- 04 April 2022

# Net benefit: using a turtle excluder device in the Adriatic Sea

Marine biologist Laura Aiudi is working on a net that saves the lives of endangered species — but still protects the livelihood of fishermen.

- [Jack Leeming](#)



Laura Aiudi is a field researcher at the University of Pisa, Italy. Credit: Elisabetta Zavoli for *Nature*

I'm from a fishing family. My grandfather was a fisherman when he was a young man, working out of Fano, the Italian town where I grew up and still live. I'm used to the smell of fish.

I'm pictured during an overnight shift on the fishing boat *RIMAS*. I work from 5 p.m. until 9 a.m. with fishermen from nearby Cesenatico on the north Adriatic Sea. It's a small boat: there's only six or so of us on board. At night, the fish are most active and we can avoid other vessels.

The nets scrape the sea bed for the catch but sometimes they also catch turtles who often die in the nets or on board. That's where I come in. The net I'm holding is designed to allow turtles to escape: it has a hole at the top they can swim out of. We call it TED — short for 'turtle excluder device'. The TED is made from a high-strength plastic, and is based on decades of work and research aimed at reducing the bycatch of turtles from trawling.

Turtles and some larger fish can leave through the escape hatch, but the current holds most of the catch in the net.

I ensure that the net is working, and that the fishermen we're collaborating with can still catch enough for their livelihoods while protecting turtles. The work is part of research by the Cetacea Foundation, based in Riccione, Italy, in collaboration with the University of Pisa, where I'm a field researcher. It is financed by the LIFE programme, the European Union's funding instrument for the environment and climate action.

I love this work. It means I'm not stuck in an office all day and instead can enjoy the ocean and work closely with people who live by the sea. I get to be a researcher who works outside, rather than being hunched over a microscope.

When my grandfather was fishing in the 1970s, there were more fish and more turtles around. At the foundation, we save 50–60 turtles a year, most of them harmed because of fishing. If we can protect turtles by rolling out this device to fishermen all across the Adriatic, I'd see this work as a success.

*Nature* **604**, 210 (2022)

*doi:* <https://doi.org/10.1038/d41586-022-00930-w>

---

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-022-00930-w>

# Research

- [\*\*An added layer of repression for human genes\*\*](#) [ 30 March 2022]  
News & Views • A protein complex called the rixosome helps to degrade RNA transcripts that linger after gene expression ceases. This discovery points to distinct roles for the rixosome in regulating chromatin in different species.
- [\*\*Cryogenic mastery aids bid to spot elusive matter creation\*\*](#) [ 06 April 2022]  
News & Views • A cubic metre of tellurium held at cryogenic temperatures over many years has enabled a search for matter created in a rare nuclear process. The feat bodes well for stabilizing other complex systems at low temperatures.
- [\*\*Genome doubling causes double trouble\*\*](#) [ 30 March 2022]  
News & Views • Human cancer cells often double their genome through an error in cell division, and this can lead to further genomic instability. A detailed analysis of the first cell cycle after genome doubling sheds light on this phenomenon.
- [\*\*Twenty-five years of nanoscale thermodynamics\*\*](#) [ 04 April 2022]  
News & Views • A paper published in 1997 brought the thermodynamics of the nineteenth century into the twenty-first century — expanding the physics of transformations involved in the operation of steam engines to the realm of molecular motors.
- [\*\*Democratizing the mapping of gene mutations to protein biophysics\*\*](#) [ 06 April 2022]  
News & Views • A general method that quantifies and disentangles the effects of a gene's mutations on the traits of its protein enables assessments of mutational effects on protein biophysics for many of the proteins of a living organism.
- [\*\*UV absorption by silicate cloud precursors in ultra-hot Jupiter WASP-178b\*\*](#) [ 06 April 2022]  
Article • Analysis is presented that indicates that there is strong evidence that SiO or Mg, the major constituents of silicate condensates, must be present in gaseous form in the atmosphere of ultra-hot Jupiter WASP-178b.
- [\*\*Search for Majorana neutrinos exploiting millikelvin cryogenics with CUORE\*\*](#) [ 06 April 2022]  
Article • The CUORE experiment finds no evidence for neutrinoless double beta decay after operating a large cryogenic TeO<sub>2</sub> calorimeter stably for several years in an extreme low-radiation environment at a temperature of 10 millikelvin.

- [Charge-density-wave-driven electronic nematicity in a kagome superconductor](#) [ 09 February 2022]  
Article • Charge-density-wave-driven electronic nematicity that occurs well below the charge-density-wave transition temperature is reported in the kagome superconductor CsV<sub>3</sub>Sb<sub>5</sub>.
- [Ultrathin ferroic HfO<sub>2</sub>–ZrO<sub>2</sub> superlattice gate stack for advanced transistors](#) [ 06 April 2022]  
Article • In the standard Si transistor gate stack, replacing conventional dielectric HfO<sub>2</sub> with an ultrathin ferroelectric–antiferroelectric HfO<sub>2</sub>–ZrO<sub>2</sub> heterostructure exhibiting the negative capacitance effect demonstrates ultrahigh capacitance without degradation in leakage and mobility, promising for ferroelectric integration into advanced logic technology.
- [Reconstructed covalent organic frameworks](#) [ 06 April 2022]  
Article • A protocol in which monomers are pre-organized using a reversible and removable urea linkage enables the production of covalent organic frameworks with higher crystallinity and porosity than those produced using standard approaches with randomly aligned monomers.
- [Autonomous fuelled directional rotation about a covalent single bond](#) [ 06 April 2022]  
Article • The molecular chemical ‘fuelling’ of the catalysis-driven motor 1-phenylpyrrole 2,2'-dicarboxylic acid, which operates by a Brownian information ratchet mechanism, facilitates dynamics that are otherwise kinetically inaccessible.
- [Multifunctional biocatalyst for conjugate reduction and reductive amination](#) [ 06 April 2022]  
Article • A biocatalytic enzyme originating from bacteria, EneIRED, facilitates amine-activated conjugate alkene reduction followed by reductive amination, efficiently preparing chiral amine diastereomers, which are commonly used in pharmaceuticals and agrochemicals.
- [Automated iterative Csp<sub>3</sub>–C bond formation](#) [ 08 February 2022]  
Article • Identification of a hyperstable boronate enables automated lego-like synthesis to access a wider range of three-dimensionally complex small organic molecules rich in Csp<sub>3</sub>–C bonds.
- [A stable hippocampal code in freely flying bats](#) [ 30 March 2022]  
Article • The hippocampal code in freely flying bats is highly stable over days and across contexts if behaviour is taken into account.
- [Entropy of city street networks linked to future spatial navigation ability](#) [ 30 March 2022]

Article • An analysis of spatial navigation in nearly 400,000 people shows, by measuring their performance in a video game, that individuals who grew up outside cities are better at navigation than those who grew up in cities.

- **Human distal lung maps and lineage hierarchies reveal a bipotent progenitor** [ 30 March 2022]

Article • Spatial transcriptomics and single-cell profiling identify previously uncharacterized cell types of human terminal and respiratory bronchioles, and show that cell differentiation and lineage trajectories are distinct from those in the mouse lung.

- **Human distal airways contain a multipotent secretory cell that can regenerate alveoli** [ 30 March 2022]

Article • Human respiratory bronchioles contain a unique population of secretory cells called respiratory airway secretory cells that are distinct from the cells in the larger proximal airways, and act as unidirectional progenitors for alveolar type 2 cells.

- **Plant phytochrome B is an asymmetric dimer with unique signalling potential** [ 30 March 2022]

Article • The inactive Pr conformer of Arabidopsis phytochrome B shows a topologically complex dimeric organization that is distinct from its prokaryotic relatives.

- **Pyrimidine inhibitors synergize with nucleoside analogues to block SARS-CoV-2** [ 07 February 2022]

Article • A combination of pyrimidine biosynthesis inhibitors and antiviral nucleoside analogues can boost the antiviral effect of nucleoside analogues against SARS-CoV-2.

- **Germinal centre-driven maturation of B cell response to mRNA vaccination** [ 15 February 2022]

Article • Sequencing of B cell receptors and expression of the corresponding monoclonal antibodies is used to characterize the evolution of the long-term B cell response to SARS-CoV-2 mRNA vaccination.

- **Genetic instability from a single S phase after whole-genome duplication** [ 30 March 2022]

Article • Extensive DNA damage occurs during the first interphase following induction of tetraploidy in human cells, largely as a result of the lower amount of protein relative to DNA.

- **Crucial role and mechanism of transcription-coupled DNA repair in bacteria** [ 30 March 2022]

Article • Integrated structure–function studies show that transcription-coupled DNA repair (TCR)—rather than global genomic repair—is responsible for most chromosomal repair events in bacteria, and that TCR mainly occurs independently of the Mfd translocase.

- **Targeting Xist with compounds that disrupt RNA structure and X inactivation** [ 30 March 2022]  
Article • A molecule identified in a screen for compounds that bind the non-coding mouse RNA Xist blocks Xist-dependent X-chromosome inactivation, demonstrating the utility of this approach for identifying drugs that target RNA.
- **Rixosomal RNA degradation contributes to silencing of Polycomb target genes** [ 30 March 2022]  
Article • The rixosome associates with Polycomb repressive complexes and chromatin and has a role in silencing of Polycomb target gene expression in human cells via degradation of nascent RNA transcripts.
- **Mapping the energetic and allosteric landscapes of protein binding domains** [ 06 April 2022]  
Article • An approach that combines deep mutational scanning with neural network-based thermodynamic modelling is used to provide comprehensive maps of the energetic and allosteric effects of mutations in two common protein domains.
- **Structure of the NLRP3 decamer bound to the cytokine release inhibitor CRID3** [ 03 February 2022]  
Article • Cryo-electron microscopy structures of human NLRP3 in its resting state and bound to the inhibitor CRID3 provide insight into the binding mechanism of CRID3 and its mode of antagonism.
- **Differential assembly diversifies GABAA receptor structures and signalling** [ 30 March 2022]  
Article • The diverse makeup and assembly of subunits augment the structure, physiology and pharmacology of GABAA receptors.
- **Structure, substrate recognition and initiation of hyaluronan synthase** [ 30 March 2022]  
Article • A cryo-electron microscopy analysis reveals how HAS selects its substrates, hydrolyses the first substrate to prime the synthesis reaction, opens a hyaluronan-conducting transmembrane channel, ensures alternating substrate polymerization and coordinates hyaluronan inside its transmembrane pore.
- **Evidence from a statewide vaccination RCT shows the limits of nudges** [ 06 April 2022]  
Matters Arising •

- NEWS AND VIEWS
- 30 March 2022

# An added layer of repression for human genes

A protein complex called the rixosome helps to degrade RNA transcripts that linger after gene expression ceases. This discovery points to distinct roles for the rixosome in regulating chromatin in different species.

- [Michael Uckelmann](#) ORCID: <http://orcid.org/0000-0002-4370-6009>  
<sup>0</sup> &
- [Chen Davidovich](#) ORCID: <http://orcid.org/0000-0002-1085-6094><sup>1</sup>

After gene expression ends, components of the gene-transcription machinery can remain in the vicinity of DNA, along with RNA transcripts. There are known pathways for transcript degradation and for gene repression in human cells — but, so far, these pathways have not been directly linked. [Writing in Nature](#), Zhou *et al.*<sup>1</sup> describe a pathway for the removal of RNA from inactive human genes that is triggered by key players of the gene-repression machinery.

## Access options

Subscribe to Nature+

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

\$29.99

monthly

[Subscribe](#)

[Subscribe to Journal](#)

Get full journal access for 1 year

\$199.00

only \$3.90 per issue

[Subscribe](#)

All prices are NET prices.

VAT will be added later in the checkout.

Tax calculation will be finalised during checkout.

[Buy article](#)

Get time limited or full article access on ReadCube.

\$32.00

[Buy](#)

All prices are NET prices.

### **Additional access options:**

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

*Nature* **604**, 41-42 (2022)

*doi:* <https://doi.org/10.1038/d41586-022-00519-3>

## **References**

1. Zhou, H. *et al.* *Nature* **604**, 167–174 (2022).
2. Jenuwein, T. & Allis, C. D. *Science* **293**, 1074–1080 (2001).

3. Blackledge, N. P. & Klose, R. J. *Nature Rev. Mol. Cell Biol.* **22**, 815–833 (2021).
4. Creamer, K. M., Kolpa, H. J. & Lawrence, J. B. *Mol. Cell* **81**, 3509–3525.e5 (2021).
5. Quinodoz, S. A. *et al.* *Cell* **184**, 5775–5790.e30 (2021).
6. Wickramasinghe, V. O. & Venkitaraman, A. R. *Mol. Cell* **61**, 496–505 (2016).
7. Li, X. & Fu, X.-D. *Nature Rev. Genet.* **20**, 503–519 (2019).
8. Fromm, L. *et al.* *Nature Commun.* **8**, 1787 (2017).
9. Shipkovenska, G., Durango, A., Kalocsay, M., Gygi, S. P. & Moazed, D. *eLife* **9**, e54341 (2020).
10. Davidovich, C., Zheng, L., Goodrich, K. J. & Cech, T. R. *Nature Struct. Mol. Biol.* **20**, 1250–1257 (2013).
11. Almeida, M., Bowness, J. S. & Brockdorff, N. *Curr. Opin. Genet. Dev.* **61**, 53–61 (2020).

---

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-022-00519-3>

- NEWS AND VIEWS
- 06 April 2022

# Cryogenic mastery aids bid to spot elusive matter creation

A cubic metre of tellurium held at cryogenic temperatures over many years has enabled a search for matter created in a rare nuclear process. The feat bodes well for stabilizing other complex systems at low temperatures.

- [Jason Detwiler](#) 0

Astrophysical observations reveal that the Universe is made almost entirely of matter, with nearly no antimatter in sight. However, laboratory and particle-collider experiments have so far observed the creation of matter and antimatter in equal parts. Big Bang theories that aim to explain the cosmic matter imbalance predict that matter could be generated without antimatter in a ‘little bang’, during an ultra-rare nuclear process called neutrinoless double- $\beta$  decay. In [a paper in Nature](#), the CUORE Collaboration<sup>1</sup> reports the most sensitive search yet for this type of decay using isotopes of tellurium. The decay was not observed, but the engineering feat was remarkable — requiring the stable operation of more than a tonne of experimental apparatus, at cryogenic temperatures close to 10 millikelvin, over several years.

## Access options

Subscribe to Nature+

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

\$29.99

monthly

[Subscribe](#)

Subscribe to Journal

Get full journal access for 1 year

\$199.00

only \$3.90 per issue

[Subscribe](#)

All prices are NET prices.

VAT will be added later in the checkout.

Tax calculation will be finalised during checkout.

Buy article

Get time limited or full article access on ReadCube.

\$32.00

[Buy](#)

All prices are NET prices.

**Additional access options:**

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

*Nature* **604**, 42-43 (2022)

*doi:* <https://doi.org/10.1038/d41586-022-00836-7>

**References**

1. The CUORE Collaboration. *Nature* **604**, 53–58 (2022).
2. Agostini, M., Benato, G., Detwiler, J., Menéndez, J. & Vissani, F. *Rev. Mod. Phys.* (in the press).
3. Ouellet, J. Preprint at <https://arxiv.org/abs/1410.1560> (2014).
4. The CUORE Collaboration. *Phys. Rev. Lett.* **124**, 122501 (2020).
5. The CUPID Interest Group. Preprint at <https://arxiv.org/abs/1907.09376> (2019).

---

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-022-00836-7>

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

- NEWS AND VIEWS
- 30 March 2022

# Genome doubling causes double trouble

Human cancer cells often double their genome through an error in cell division, and this can lead to further genomic instability. A detailed analysis of the first cell cycle after genome doubling sheds light on this phenomenon.

- [Yonatan Eliezer](#)<sup>0</sup> &
- [Uri Ben-David](#)<sup>1</sup>

Understanding how genomic abnormalities arise in cancer cells might offer new ways of developing treatments. [Writing in Nature](#), Gemble *et al.*<sup>1</sup> present an examination of the effects of genome doubling — an event that can lead to tumour formation.

## Access options

Subscribe to Nature+

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

\$29.99

monthly

[Subscribe](#)

Subscribe to Journal

Get full journal access for 1 year

\$199.00

only \$3.90 per issue

[Subscribe](#)

All prices are NET prices.

VAT will be added later in the checkout.

Tax calculation will be finalised during checkout.

Buy article

Get time limited or full article access on ReadCube.

\$32.00

[Buy](#)

All prices are NET prices.

### **Additional access options:**

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

*Nature* **604**, 44-45 (2022)

doi: <https://doi.org/10.1038/d41586-022-00849-2>

## **References**

1. Gemble, S. *et al.* *Nature* **604**, 146–151 (2022).
2. Zack, T. I. *et al.* *Nature Genet.* **45**, 1134–1140 (2013).
3. Bielski, C. M. *et al.* *Nature Genet.* **50**, 1189–1195 (2018).

4. Dewhurst, S. M. *et al.* *Cancer Discov.* **4**, 175–185 (2014).
5. López, S. *et al.* *Nature Genet.* **52**, 283–293 (2020).
6. Quinton, R. J. *et al.* *Nature* **590**, 492–497 (2021).
7. Prasad, K. *et al.* *Cancer Res.* <https://doi.org/10.1158/0008-5472.CAN-21-2065> (2022).
8. Ganem, N. J., Storchova, Z. & Pellman, D. *Curr. Opin. Genet. Dev.* **17**, 157–162 (2007).
9. Nicholson, J. M. & Cimini, D. *Adv. Cancer Res.* **112**, 43–75 (2011).
10. Storchová, Z. *et al.* *Nature* **443**, 541–547 (2006).
11. Fujiwara, T. *et al.* *Nature* **437**, 1043–1047 (2005).
12. Kuznetsova, A. Y. *et al.* *Cell Cycle* **14**, 2810–2820 (2015).
13. Selmecki, A. M. *et al.* *Nature* **519**, 349–352 (2015).
14. Wangsa, D. *et al.* *FASEB J.* **32**, 3502–3517 (2018).
15. Kuffer, C., Kuznetsova, A. Y. & Storchová, Z. *Chromosoma* **122**, 305–318 (2013).
16. Bester, A. C. *et al.* *Cell* **145**, 435–446 (2011).
17. DeGregori, J., Kowalik, T. & Nevins, J. R. *Mol. Cell. Biol.* **15**, 4215–4224 (1995).
18. Stephens, P. J. *et al.* *Cell* **144**, 27–40 (2011).
19. Viganó, C. *et al.* *Mol. Biol. Cell* **29**, 1031–1047 (2018).

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

- NEWS AND VIEWS
- 04 April 2022

# Twenty-five years of nanoscale thermodynamics

A paper published in 1997 brought the thermodynamics of the nineteenth century into the twenty-first century — expanding the physics of transformations involved in the operation of steam engines to the realm of molecular motors.

- [Chase P. Broedersz](#)<sup>0</sup> &
- [Pierre Ronceray](#)<sup>1</sup>

The history of thermodynamics begins in 1824, when the French engineer Nicolas Léonard Sadi Carnot observed that a steam engine's efficiency has a natural limit. The German physicist Rudolf Clausius made sense of this with a mathematical inequality that became known as the second law of thermodynamics: a quantity that he called entropy keeps increasing until a system reaches equilibrium. This law has since become a fixed feature of university physics curricula, and remained unimproved for more than a century — until 25 years ago, when Christopher Jarzynski<sup>1</sup> [proposed an improved relationship](#) that has intriguing consequences for microscopic systems, such as the tiny engines that drive living organisms.

## Access options

Subscribe to Nature+

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

\$29.99

monthly

[Subscribe](#)

Subscribe to Journal

Get full journal access for 1 year

\$199.00

only \$3.90 per issue

[Subscribe](#)

All prices are NET prices.

VAT will be added later in the checkout.

Tax calculation will be finalised during checkout.

Buy article

Get time limited or full article access on ReadCube.

\$32.00

[Buy](#)

All prices are NET prices.

**Additional access options:**

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

*Nature* **604**, 46-47 (2022)

*doi:* <https://doi.org/10.1038/d41586-022-00869-y>

**References**

1. Jarzynski, C. *Phys. Rev. Lett.* **78**, 2690–2693 (1997).
2. Liphardt, J., Dumont, S., Smith, S. B., Tinoco, I. Jr, & Bustamante, C. *Science*, **296**, 1832–1835 (2002).
3. Pietzonka, P., Barato, A. C. & Seifert, U. *J. Stat. Mech.* **2016**, 124004 (2016).
4. Evans, D. J., Cohen, E. G. D. & Morriss, G. P. *Phys. Rev. Lett.* **71**, 2401–2404 (1993).
5. Gallavotti, G. & Cohen, E. G. D. *Phys. Rev. Lett.* **74**, 2694–2697 (1995).
6. Crooks, G. E. *Phys. Rev. E* **60**, 2721–2726 (1999).
7. Seifert, U. *Rep. Prog. Phys.* **75**, 126001 (2012).
8. Saira, O.-P. *et al.* *Phys. Rev. Lett.* **109**, 180601 (2012).
9. Park, S. & Schulten, K. *J. Chem. Phys.* **120**, 5946–5961 (2004).
10. Barato, A. C. & Seifert, U. *Phys. Rev. Lett.* **114**, 158101 (2015).
11. Gingrich, T. R., Horowitz, J. M., Perunov, N. & England, J. L. *Phys. Rev. Lett.* **116**, 120601 (2016).

---

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-022-00869-y>

- NEWS AND VIEWS
- 06 April 2022

# Democratizing the mapping of gene mutations to protein biophysics

A general method that quantifies and disentangles the effects of a gene's mutations on the traits of its protein enables assessments of mutational effects on protein biophysics for many of the proteins of a living organism.

- [Debora S. Marks](#) <sup>0</sup> &
- [Stephen W. Michnick](#) <sup>1</sup>

Proteins function through their interactions with other biomolecules, and can be modulated not only by changes to the amino-acid residues at the protein–biomolecule interface, but also by changes at distal sites — a phenomenon called allostery. Discovering which protein sites are allosteric has been difficult, because methods that simply and comprehensively quantify allosteric effects have been lacking. [Writing in Nature](#), Faure *et al.*<sup>1</sup> report the first integrated method to globally map, quantify and distinguish the effects of mutations on allostery and protein stability. Their approach is also easy to do, making it as accessible to geneticists as it is to biophysicists. This method will have broad applications in biotechnology and drug discovery.

## Access options

Subscribe to Nature+

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

\$29.99

monthly

[Subscribe](#)

Subscribe to Journal

Get full journal access for 1 year

\$199.00

only \$3.90 per issue

[Subscribe](#)

All prices are NET prices.

VAT will be added later in the checkout.

Tax calculation will be finalised during checkout.

Buy article

Get time limited or full article access on ReadCube.

\$32.00

[Buy](#)

All prices are NET prices.

**Additional access options:**

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

*Nature* **604**, 47-48 (2022)

doi: <https://doi.org/10.1038/d41586-022-00870-5>

**References**

1. Faure, A. J. *et al.* *Nature* **604**, 175–183 (2022).
2. Fowler, D. M. & Fields, S. *Nature Methods* **11**, 801–807 (2014).
3. Rubin, A. F. *et al.* Preprint at bioRxiv  
<https://doi.org/10.1101/2021.11.29.470445> (2021).
4. Starr, T. N. *et al.* *Cell* **182**, 1295–1310 (2020).
5. Greaney, A. J. *et al.* *Cell Host Microbe* **29**, 463–476 (2021).
6. Hopf, T. A. *et al.* *Nature Biotechnol.* **35**, 128–135 (2017).
7. Frazer, J. *et al.* *Nature* **599**, 91–95 (2021).
8. Pelletier, J. N., Campbell-Valois, F.-X. & Michnick, S. W. *Proc. Natl Acad. Sci. USA* **95**, 12141–12146 (1998).
9. Campbell-Valois, F.-X., Tarassov, K. & Michnick, S. W. *Proc. Natl Acad. Sci. USA* **102**, 14988–14993 (2005).
10. Hein, M. Y. *et al.* *Cell* **163**, 712–723 (2015).

---

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-022-00870-5>

- Article
- [Published: 06 April 2022](#)

# UV absorption by silicate cloud precursors in ultra-hot Jupiter WASP-178b

- [Joshua D. Lothringer](#) ORCID: [orcid.org/0000-0003-3667-8633<sup>1,2 na1</sup>](https://orcid.org/0000-0003-3667-8633),
- [David K. Sing](#) ORCID: [orcid.org/0000-0001-6050-7645<sup>2,3 na1</sup>](https://orcid.org/0000-0001-6050-7645),
- [Zafar Rustamkulov](#)<sup>3</sup>,
- [Hannah R. Wakeford](#)<sup>4</sup>,
- [Kevin B. Stevenson](#) ORCID: [orcid.org/0000-0002-7352-7941<sup>3,5</sup>](https://orcid.org/0000-0002-7352-7941),
- [Nikolay Nikolov](#) ORCID: [orcid.org/0000-0002-6500-3574<sup>6</sup>](https://orcid.org/0000-0002-6500-3574),
- [Panayotis Lavvas](#) ORCID: [orcid.org/0000-0002-5360-3660<sup>7</sup>](https://orcid.org/0000-0002-5360-3660),
- [Jessica J. Spake](#)<sup>8</sup> &
- [Autumn T. Winch](#)<sup>9</sup>

*Nature* volume **604**, pages 49–52 (2022)

- 683 Accesses
- 500 Altmetric
- [Metrics details](#)

## Subjects

- [Atmospheric chemistry](#)
- [Exoplanets](#)
- [Giant planets](#)

# Abstract

Aerosols have been found to be nearly ubiquitous in substellar atmospheres<sup>1,2,3</sup>. The precise temperature at which these aerosols begin to form in exoplanets has yet to be observationally constrained. Theoretical models and observations of muted spectral features indicate that silicate clouds play an important role in exoplanets between at least 950 and 2,100 K (ref. <sup>4</sup>). Some giant planets, however, are thought to be hot enough to avoid condensation altogether<sup>5,6</sup>. Here we report the near-ultraviolet transmission spectrum of the ultra-hot Jupiter WASP-178b (approximately 2,450 K), which exhibits substantial absorption. Bayesian retrievals indicate the presence of gaseous refractory species containing silicon and magnesium, which are the precursors to condensate clouds at lower temperatures. SiO, in particular, has not previously, to our knowledge, been detected in exoplanets, but the presence of SiO in WASP-178b is consistent with theoretical expectations as the dominant Si-bearing species at high temperatures. These observations allow us to re-interpret previous observations of HAT-P-41b and WASP-121b that did not consider SiO, to suggest that silicate cloud formation begins on exoplanets with equilibrium temperatures between 1,950 and 2,450 K.

This is a preview of subscription content

## Access options

Subscribe to Nature+

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

\$29.99

monthly

[Subscribe](#)

Subscribe to Journal

Get full journal access for 1 year

\$199.00

only \$3.90 per issue

[Subscribe](#)

All prices are NET prices.

VAT will be added later in the checkout.

Tax calculation will be finalised during checkout.

Buy article

Get time limited or full article access on ReadCube.

\$32.00

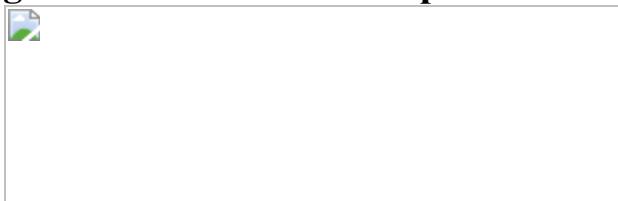
[Buy](#)

All prices are NET prices.

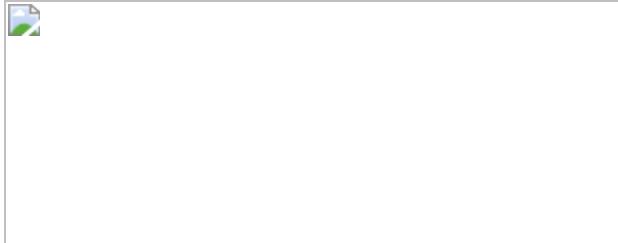
### **Additional access options:**

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

**Fig. 1: WASP-178b NUV-optical transmission spectrum.**



**Fig. 2: Comparison of NUV-optical transmission spectra.**



**Fig. 3: Atmospheric structures and condensation curves.**



## Data availability

The raw data from this study, HST Program 16068, is publicly available via the Space Science Telescope Institute's Mikulski Archive for Space Telescopes (<https://archive.stsci.edu/>).

## Code availability

The raw data was reduced with the available STScI CALWF3 pipeline and spectra were extracted with the public IRAF apall routines. The light curve fitting used custom routines that we opt not to make public due to undocumented intricacies. Model and retrievals were generated using PHOENIX, which is a proprietary code but described in many publications, for example, refs. [66,67](#).

## References

1. Cushing, M. C. et al. A Spitzer infrared spectrograph spectral sequence of M, L, and T dwarfs. *Astrophys. J.* **648**, 614–628 (2006).

2. Saumon, D. & Marley, M. S. The evolution of L and T dwarfs in color-magnitude diagrams. *Astrophys. J.* **689**, 1327–1344 (2008).
3. Burningham, B. et al. Cloud busting: enstatite and quartz clouds in the atmosphere of 2M2224-0158. *Mon. Not. R. Astron. Soc.* **506**, 1944–1961 (2021).
4. Gao, P. et al. Aerosol composition of hot giant exoplanets dominated by silicates and hydrocarbon hazes. *Nature Astron.* **4**, 951–956 (2020).
5. Lothringer, J. D. et al. An HST/STIS optical transmission spectrum of warm Neptune GJ 436b. *Astron. J.* **155**, 66 (2018).
6. Kitzmann, D. et al. The peculiar atmospheric chemistry of KELT-9b. *Astrophys. J.* **863**, 183 (2018).
7. Hellier, C. et al. WASP-South hot Jupiters: WASP-178b, WASP-184b, WASP-185b, and WASP-192b. *Mon. Not. R. Astron. Soc.* **490**, 1479–1487 (2019).
8. Rodríguez Martínez, R. et al. KELT-25 b and KELT-26 b: a hot Jupiter and a substellar companion transiting young stars observed by TESS. *Astron. J.* **160**, 111 (2020).
9. Matsushima, S. Radiative opacity in stellar atmospheres. II. Effect of ultraviolet continuum on the photospheric radiation field. *Astrophys. J.* **154**, 715 (1968).
10. Fontenla, J. M., Stancil, P. C. & Landi, E. Solar spectral irradiance, solar activity, and the near-ultra-violet. *Astrophys. J.* **809**, 157 (2015).
11. Sharp, C. M. & Burrows, A. Atomic and molecular opacities for brown dwarf and giant planet atmospheres. *Astrophys. J. Supp.* **168**, 140–166 (2007).
12. Lothringer, J. D., Fu, G., Sing, D. K. & Barman, T. S. UV exoplanet transmission spectral features as probes of metals and rainout. *Astrophys. J. Lett.* **898**, L14 (2020).

13. Hoeijmakers, H. J. et al. Hot exoplanet atmospheres resolved with transit spectroscopy (HEARTS). IV. A spectral inventory of atoms and molecules in the high-resolution transmission spectrum of WASP- 121 b. *Astron. Astrophys.* **641**, A123 (2020).
14. Stangret, M. et al. Detection of Fe I and Fe II in the atmosphere of MASCARA-2b using a cross- correlation method. *Astron. Astrophys.* **638**, A26 (2020).
15. Ehrenreich, D. et al. Nightside condensation of iron in an ultrahot giant exoplanet. *Nature* **580**, 597–601 (2020).
16. Kesseli, A. Y. & Snellen, I. A. G. Confirmation of asymmetric iron absorption in WASP-76b with HARPS. *Astrophys. J. Lett.* **908**, L17 (2021).
17. Sing, D. K. et al. The Hubble Space Telescope PanCET program: exospheric Mg II and Fe II in the near-ultraviolet transmission spectrum of WASP-121b using jitter decorrelation. *Astron. J.* **158**, 91 (2019).
18. Gibson, N. P. et al. Detection of Fe I in the atmosphere of the ultra-hot Jupiter WASP-121b, and a new likelihood-based approach for Doppler-resolved spectroscopy. *Mon. Not. R. Astron. Soc.* **493**, 2215–2228 (2020).
19. Cabot, S. H. C., Madhusudhan, N., Welbanks, L., Piette, A. & Gandhi, S. Detection of neutral atomic species in the ultra-hot Jupiter WASP-121b. *Mon. Not. R. Astron. Soc.* **494**, 363–377 (2020).
20. Hoeijmakers, H. J. et al. A spectral survey of an ultra-hot Jupiter. Detection of metals in the transmission spectrum of KELT-9 b. *Astron. Astrophys.* **627**, A165 (2019).
21. Merritt, S. R. et al. An inventory of atomic species in the atmosphere of WASP-121b using UVES high-resolution spectroscopy. *Mon. Not. R. Astron. Soc.* **506**, 3853–3871 (2021).

22. Wakeford, H. R. et al. Into the UV: a precise transmission spectrum of HAT-P-41b using Hubble’s WFC3/UVIS G280 grism. *Astron. J.* **159**, 204 (2020).
23. Visscher, C., Lodders, K. & Fegley, J. B. Atmospheric chemistry in giant planets, brown dwarfs, and low-mass dwarf stars. III. Iron, magnesium, and silicon. *Astrophys. J.* **716**, 1060–1075 (2010).
24. Parmentier, V., Showman, A. P. & Fortney, J. J. The cloudy shape of hot Jupiter thermal phase curves. *Mon. Not. R. Astron. Soc.* **501**, 78–108 (2021).
25. Roman, M. T. et al. Clouds in three-dimensional models of hot Jupiters over a wide range of temperatures. I. Thermal structures and broadband phase-curve predictions. *Astrophys. J.* **908**, 101 (2021).
26. Helling, C. et al. Cloud property trends in hot and ultra-hot giant gas planets (WASP-43b, WASP-103b, WASP-121b, HAT-P-7b, and WASP-18b). *Astron. Astrophys.* **649**, A44 (2021).
27. Thorngren, D., Gao, P. & Fortney, J. J. The intrinsic temperature and radiative–convective boundary depth in the atmospheres of hot Jupiters. *Astrophys. J. Lett.* **884**, L6 (2019).
28. Hörst, S. M. et al. Haze production rates in super-Earth and mini-Neptune atmosphere experiments. *Nature Astron.* **2**, 303–306 (2018).
29. Fleury, B., Gudipati, M. S., Henderson, B. L. & Swain, M. Photochemistry in hot H<sub>2</sub>-dominated exoplanet atmospheres. *Astrophys. J.* **871**, 158 (2019).
30. Kempton, E. M. R. et al. A framework for prioritizing the TESS planetary candidates most amenable to atmospheric characterization. *Publ. Astron. Soc. Pac.* **130**, 114401 (2018).
31. Mullally, S. E., Rodriguez, D. R., Stevenson, K. B. & Wakeford, H. R. The Exo.MAST table for JWST exoplanet atmosphere observability. *Res. Notes AAS* **3**, 193 (2019).

32. Luna, J. L. & Morley, C. V. Empirically determining substellar cloud compositions in the era of the James Webb Space Telescope. *Astrophys. J.* **920**, 146 (2021).
33. Evans, T. M. et al. An optical transmission spectrum for the ultra-hot Jupiter WASP-121b measured with the Hubble Space Telescope. *Astron. J.* **156**, 283 (2018).
34. Sing, D. K. et al. A continuum from clear to cloudy hot-Jupiter exoplanets without primordial water depletion. *Nature* **529**, 59–62 (2016).
35. van Dokkum, P. G. Cosmic-ray rejection by Laplacian edge detection. *Publ. Astron. Soc. Pac.* **113**, 1420–1427 (2001).
36. Pirzkal, N., Hilbert, B. & Rothberg, B. *Trace and Wavelength Calibrations of the UVIS G280 +1/-1 Grism Orders* Space Telescope WFC Instrument Science Report (Space Telescope Science Institute, 2017).
37. Mandel, K. & Agol, E. Analytic light curves for planetary transit searches. *Astrophys. J. Lett.* **580**, L171–L175 (2002).
38. Pont, F., Zucker, S. & Queloz, D. The effect of red noise on planetary transit detection. *Mon. Not. R. Astron. Soc.* **373**, 231–242 (2006).
39. Winn, J. N. et al. The Transit Light Curve Project. VII. The not-so-bloated exoplanet HAT-P-1b. *Astron. J.* **134**, 1707–1712 (2007).
40. Hauschildt, P. H., Allard, F. & Baron, E. The NextGen Model Atmosphere Grid for  $3000 \leq \text{Teff} \leq 10,000$  K. *Astrophys. J.* **512**, 377–385 (1999).
41. Sing, D. K. Stellar limb-darkening coefficients for CoRot and Kepler. *Astron. Astrophys.* **510**, A21 (2010).
42. Schaller, G., Schaerer, D., Meynet, G. & Maeder, A. New grids of stellar models from 0.8 to 120 M solar at Z=0.020 and Z=0.001.

*Astron. Astrophys. Suppl. Ser.* **96**, 269 (1992).

43. Barman, T. S., Hauschildt, P. H. & Allard, F. Irradiated planets. *Astrophys. J.* **556**, 885–895 (2001).
44. Lothringer, J. D., Barman, T. & Koskinen, T. Extremely irradiated hot Jupiters: non-oxide inversions, H<sup>-</sup> opacity, and thermal dissociation of molecules. *Astrophys. J.* **866**, 27 (2018).
45. Lothringer, J. D. & Barman, T. The influence of host star spectral type on ultra-hot Jupiter atmospheres. *Astrophys. J.* **876**, 69 (2019).
46. Hubeny, I., Burrows, A. & Sudarsky, D. A possible bifurcation in atmospheres of strongly irradiated stars and planets. *Astrophys. J.* **594**, 1011–1018 (2003).
47. Fortney, J. J., Lodders, K., Marley, M. S. & Freedman, R. S. A unified theory for the atmospheres of the hot and very hot Jupiters: two classes of irradiated atmospheres. *Astrophys. J.* **678**, 1419–1435 (2008).
48. Diamond-Lowe, H., Stevenson, K. B., Bean, J. L., Line, M. R. & Fortney, J. J. New analysis indicates no thermal inversion in the atmosphere of HD 209458b. *Astrophys. J.* **796**, 66 (2014).
49. Lewis, N. K. et al. Into the UV: the atmosphere of the hot Jupiter HAT-P-41b revealed. *Astrophys. J. Lett.* **902**, L19 (2020).
50. Lothringer, J. D. & Barman, T. S. The PHOENIX exoplanet retrieval algorithm and using H<sup>-</sup> opacity as a probe in ultrahot Jupiters. *Astron. J.* **159**, 289 (2020).
51. ter Braak, C. J. F. & Vrugt, J. A. Differential evolution markov chain with snooker updater and fewer chains. *Stat. Comput.* **18**, 435–446 (2008).
52. Lothringer, J. D. et al. A new window into planet formation and migration: refractory-to-volatile elemental ratios in ultra-hot Jupiters. *Astrophys. J.* **914**, 12 (2021).

53. Wilson, J. et al. Gemini/GMOS optical transmission spectroscopy of WASP-121b: signs of variability in an ultra-hot Jupiter? *Mon. Not. R. Astron. Soc.* **503**, 4787–4801 (2021).
54. Parmentier, V. & Guillot, T. A non-grey analytical model for irradiated atmospheres. I. Derivation. *Astron. Astrophys.* **562**, A133 (2014).
55. Gelman, A. & Rubin, D. B. Inference from iterative simulation using multiple sequences. *AStat. Sci.* **7**, 457–511 (1992).
56. MacDonald, R. J. & Madhusudhan, N. HD 209458b in new light: detection of nitrogen chemistry, patchy clouds and sub-solar water. *Mon. Not. R. Astron. Soc.* **469**, 1979–1996 (2017).
57. McCullough, P. R., Crouzet, N., Deming, D. & Madhusudhan, N. water vapor in the spectrum of the extrasolar planet HD 189733b. I. The transit. *Astrophys. J.* **791**, 55 (2014).
58. Rackham, B. V., Apai, D. & Giampapa, M. S. The transit light source effect: false spectral features and incorrect densities for M-dwarf transiting planets. *Astrophys. J.* **853**, 122 (2018).
59. Rackham, B. V., Apai, D. & Giampapa, M. S. The transit light source effect. II. The impact of stellar heterogeneity on transmission spectra of planets orbiting broadly Sun-like stars. *Astron. J.* **157**, 96 (2019).
60. Kirk, J. et al. ACCESS and LRG-BEASTS: a precise new optical transmission spectrum of the ultrahot Jupiter WASP-103b. *Astron. J.* **162**, 34 (2021).
61. Kochanek, C. S. et al. The All-Sky Automated Survey for Supernovae (ASAS-SN) Light Curve Server v1.0. *Publ. Astron. Soc. Pac.* **129**, 104502 (2017).
62. Jayasinghe, T. et al. The ASAS-SN catalogue of variable stars – II. Uniform classification of 412 000 known variables. *Mon. Not. R. Astron. Soc.* **486**, 1907–1943 (2019).

63. Lecavelier Des Etangs, A., Pont, F., Vidal-Madjar, A. & Sing, D. Rayleigh scattering in the transit spectrum of HD 189733b. *Astron. Astrophys.* **481**, L83–L86 (2008).
64. Ohno, K. & Kawashima, Y. Super-Rayleigh slopes in transmission spectra of exoplanets generated by photochemical haze. *Astrophys. J. Lett.* **895**, L47 (2020).
65. Powell, D. et al. Transit signatures of inhomogeneous clouds on hot Jupiters: insights from micro- physical cloud modeling. *Astrophys. J.* **887**, 170 (2019).
66. Espinoza, N. & Jones, K. Constraining mornings and evenings on distant worlds: a new semianalytical approach and prospects with transmission spectroscopy. *Astron. J.* **162**, 165 (2021).
67. Mikal-Evans, T. et al. Diurnal variations in the stratosphere of an ultrahot planet. *Nat. Astron.* <https://doi.org/10.1038/s41550-021-01592-w> (2021).
68. Showman, A. P., Fortney, J. J., Lewis, N. K. & Shabram, M. Doppler signatures of the atmospheric circulation on hot Jupiters. *Astrophys. J.* **762**, 24 (2013).
69. Woitke, P. et al. Equilibrium chemistry down to 100 K. Impact of silicates and phyllosilicates on the carbon to oxygen ratio. *Astron. Astrophys.* **614**, A1 (2018).

## Acknowledgements

We thank the UV-SCOPE team for relevant discussions. We thank T. Barman for the use of the computing resources used in the calculation of the atmospheric retrievals. Support for this work was provided by NASA through grant number HST-GO-16086 from the Space Telescope Science Institute, which is operated by AURA, Inc., under NASA contract NAS 5-26555. This research has made use of the NASA Astrophysics Data System and the NASA Exoplanet Archive, which is operated by the California

Institute of Technology, under contract with the National Aeronautics and Space Administration under the Exoplanet Exploration Program.

## Author information

### Author notes

1. These authors contributed equally: Joshua D. Lothringer, David K. Sing

## Affiliations

1. Department of Physics, Utah Valley University, Orem, UT, USA  
Joshua D. Lothringer
2. Department of Physics and Astronomy, Johns Hopkins University, Baltimore, MD, USA  
Joshua D. Lothringer & David K. Sing
3. Department of Earth & Planetary Sciences, Johns Hopkins University, Baltimore, MD, USA  
David K. Sing, Zafar Rustamkulov & Kevin B. Stevenson
4. School of Physics, University of Bristol, HH Wills Physics Laboratory, Bristol, UK  
Hannah R. Wakeford
5. Applied Physics Laboratory, Johns Hopkins University, Laurel, MD, USA  
Kevin B. Stevenson
6. Space Telescope Science Institute, Baltimore, MD, USA

Nikolay Nikolov

7. Groupe de Spectrométrie Moléculaire et Atmosphérique, Université de Reims, Champagne-Ardenne, CNRS UMR F-7331, Reims, France

Panayotis Lavvas

8. Department of Astronomy, California Institute of Technology,  
Pasadena, CA, USA

Jessica J. Speke

9. Department of Physics, Bryn Mawr College, Bryn Mawr, PA, USA

Autumn T. Winch

## Contributions

J.D.L. and D.K.S. contributed equally to this work. J.D.L. led the observing proposal with the assistance of D.K.S., Z.R., H.R.W., K.B.S., N.N. and P.L. J.D.L. also led the retrieval analysis. D.K.S. led the data analysis with contributions from Z.R., H.R.W., J.J.S. and A.T.W. All authors discussed the data analysis and interpretation and commented on the manuscript.

## Corresponding authors

Correspondence to [Joshua D. Lothringer](#) or [David K. Sing](#).

## Ethics declarations

## Competing interests

The authors declare no competing interests.

## Peer review

## Peer review information

*Nature* thanks Tad Komacek and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

## Additional information

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## Extended data figures and tables

### [Extended Data Fig. 1 White light curves of the WASP-178b HST/WFC3/UVISG280 transit.](#)

Error bars show the  $1-\sigma$  uncertainties. The left column (**a, c, e**) shows the +1 spectral order, while the right column (**b, d, f**) shows the -1 spectral order. The top row (**a, b**) are the raw light curves, the middle row (**c, d**) are the light curves with systematics removed and a transit fit, and the bottom row (**e, f**) are the residuals with the standard deviation of the residuals also shown (dotted lines). Plots of the binned residual RMS are also shown as insets.

### [Extended Data Fig. 2 WASP-178b NUV-optical light curve comparison.](#)

Two example fitted light curves with  $1-\sigma$  uncertainties from the +1 spectral order from spectroscopic bins covering 0.2412 (**a, c, e**) and 0.5875  $\mu\text{m}$  (**b, d, f**), with transit depths of  $1.48 \pm 0.04\%$  and  $1.16 \pm 0.03\%$ , respectively. The rows are the same as in Extended Data Fig. 1.

### [Extended Data Fig. 3 WASP-178b spectral order comparison.](#)

WFC3/UVIS G280 transmission spectrum of WASP-178b (with  $1-\sigma$  uncertainties) from the +1 (blue) and -1 order (red). The -1 order shows larger uncertainties due to a reduced throughput, but the transmission spectra show good agreement including an enhanced NUV absorption between 0.2 and 0.3  $\mu\text{m}$ .

### **Extended Data Fig. 4 WASP-178b transit asymmetry analysis.**

**a**, The 0.18–0.28  $\mu\text{m}$  NUV light curve of WASP-178 b (with  $1-\sigma$  uncertainties), with the best-fitting symmetric light curve, and an asymmetric light curve representing a scenario with a hotter/larger trailing terminator, and a colder/smaller leading terminator. The radius of the leading terminator was set to the optical value, and the trailing terminator was fixed to the value that fits the NUV transit depth. The inset shows the RMS scatter of the residuals as a function of number of points per bin, N. **b**, Residuals to the symmetric and asymmetric light curve fits.

### **Extended Data Fig. 5 WASP-178b atmospheric retrieval posterior distribution.**

2-D cross- sections of the retrieved posterior distribution with 1-D marginalized distribution for the fitted parameters. The quoted quantities are the mean and  $1-\sigma$  retrieved values. The first five parameters are the temperature structure parameterization from ref. <sup>57</sup>, the sixth is the reference radius, and the final eight are the various atomic and molecular abundances.

### **Extended Data Fig. 6 WASP-178b NUV-optical transmission spectrum (no SiO).**

Same as Fig. 1, but for the retrieval without SiO. Note the combined ability of Mg I and Fe II absorption to generate the large short-wavelength transit depths.

### **Extended Data Fig. 7 Chemical equilibrium of Si and Fe.**

Partial pressures of important silicon- bearing species (**a**) and iron-bearing species (**b**) at 1 mbar as a function of temperature. Equilibrium chemical abundances were calculated using GGchem<sup>69</sup>.

## **Extended Data Fig. 8 High-resolution HST/STIS/E230M transmission spectrum of WASP- 178b.**

NUV high-resolution transit spectra of WASP-178b (with  $1-\sigma$  uncertainties) compared to WASP-121b around the Fe II (**a**) and Mg II (**b**) lines. Shown are the spectra from STIS E230M for WASP-178b (red), WASP-121b<sup>17</sup> (grey), and the low resolution UVIS spectra (blue). While WASP-121b shows strong Fe II and Mg II absorption features, the WASP-178b E230M spectra is consistent with the broadband NUV continuum with no Fe II or Mg II.

**Extended Data Table 1 WASP-178 HST/WFC3/UVIS transmission spectrum and noise properties**

**Extended Data Table 2 WASP-178b fitted and retrieved orbital and atmosphere parameters**

## **Supplementary information**

### **Peer Review File**

## **Rights and permissions**

### **Reprints and Permissions**

## **About this article**

### **Cite this article**

Lothringer, J.D., Sing, D.K., Rustamkulov, Z. *et al.* UV absorption by silicate cloud precursors in ultra-hot Jupiter WASP-178b. *Nature* **604**, 49–52 (2022). <https://doi.org/10.1038/s41586-022-04453-2>

- Received: 20 September 2021
- Accepted: 24 January 2022
- Published: 06 April 2022
- Issue Date: 07 April 2022
- DOI: <https://doi.org/10.1038/s41586-022-04453-2>

## Share this article

Anyone you share the following link with will be able to read this content:

[Get shareable link](#)

Sorry, a shareable link is not currently available for this article.

[Copy to clipboard](#)

Provided by the Springer Nature SharedIt content-sharing initiative

---

This article was downloaded by **calibre** from <https://www.nature.com/articles/s41586-022-04453-2>

[Download PDF](#)

- Article
- Open Access
- [Published: 06 April 2022](#)

# Search for Majorana neutrinos exploiting millikelvin cryogenics with CUORE

- [The CUORE Collaboration](#)

*Nature* volume **604**, pages 53–58 (2022)

- 1057 Accesses
- 75 Altmetric
- [Metrics details](#)

## Subjects

- [Experimental nuclear physics](#)
- [Experimental particle physics](#)

## Abstract

The possibility that neutrinos may be their own antiparticles, unique among the known fundamental particles, arises from the symmetric theory of fermions proposed by Ettore Majorana in 1937<sup>1</sup>. Given the profound consequences of such Majorana neutrinos, among which is a potential explanation for the matter–antimatter asymmetry of the universe via leptogenesis<sup>2</sup>, the Majorana nature of neutrinos commands intense experimental scrutiny globally; one of the primary experimental probes is neutrinoless double beta ( $0\nu\beta\beta$ ) decay. Here we show results from the search for  $0\nu\beta\beta$  decay of  $^{130}\text{Te}$ , using the latest advanced cryogenic calorimeters with the CUORE experiment<sup>3</sup>. CUORE, operating just 10 millikelvin above absolute zero, has pushed the state of the art on three frontiers: the sheer mass held at such ultralow

temperatures, operational longevity, and the low levels of ionizing radiation emanating from the cryogenic infrastructure. We find no evidence for  $0\nu\beta\beta$  decay and set a lower bound of the process half-life as  $2.2 \times 10^{25}$  years at a 90 per cent credibility interval. We discuss potential applications of the advances made with CUORE to other fields such as direct dark matter, neutrino and nuclear physics searches and large-scale quantum computing, which can benefit from sustained operation of large payloads in a low-radioactivity, ultralow-temperature cryogenic environment.

## Main

The standard model of particle physics is a successful paradigm for the number, properties and interactions of fundamental particles. Nevertheless, the observation of neutrino oscillations indicates the incompleteness of the standard model: they imply non-vanishing neutrino masses, requiring an extension of the standard model, and violate three accidental symmetries connected to the flavour lepton numbers  $L_e$ ,  $L_\mu$  and  $L_\tau$ , leaving the difference between the baryon and lepton number,  $B - L$ , as the only unprobed quantity. A promising process to experimentally test  $B - L$  is neutrinoless double beta ( $0\nu\beta\beta$ ) decay, in which a nucleus of mass number  $A$  and charge  $Z$  decays by the emission of only two electrons:  $(A, Z) \rightarrow (A, Z + 2) + 2e^-$ . We highlight that this process creates two electrons, namely two matter particles<sup>4</sup>. This decay can be mediated by various non-standard model mechanisms involving Majorana neutrino masses. A minimal extension of the standard model Lagrangian adds heavy Majorana neutrinos that mix with the known neutrinos to produce a set of light Majorana neutrinos, explaining the observed light neutrino masses<sup>5</sup> and at the same time providing a mechanism to explain the baryon asymmetry in the universe<sup>2</sup>. At this time, experimental searches for  $0\nu\beta\beta$  decay are the most sensitive means to corroborate this framework.

The  $0\nu\beta\beta$  decay signature is a peak in the spectrum of summed energy of the two emitted electrons at the mass difference ( $Q_{\beta\beta}$ ) between the parent and daughter nuclei. A worldwide quest is ongoing, involving a range of nuclei such as  $^{76}\text{Ge}$ <sup>6,7</sup>,  $^{136}\text{Xe}$ <sup>8,9</sup> and  $^{130}\text{Te}$ . The latter, in the form of  $\text{TeO}_2$  cryogenic calorimeters, is used by the Cryogenic Underground Observatory for Rare Events, CUORE<sup>10,11</sup>.

To fully exploit the potential of  $\text{TeO}_2$  crystals as cryogenic calorimeters, the CUORE Collaboration designed and built to our knowledge the largest dilution refrigerator ever constructed, capable of cooling approximately 1.5 t of material to a temperature of approximately 10 mK and maintaining it for years with a 90% duty cycle (1 t = 1,000 kg). In this Article, we describe the performance of CUORE over a four-year

measurement campaign and the results of a new high-sensitivity  $0\nu\beta\beta$  decay search with over 1 t yr of  $\text{TeO}_2$  exposure.

## The CUORE experiment

CUORE is the culmination of thirty years of  $0\nu\beta\beta$  decay searches with  $\text{TeO}_2$  cryogenic calorimeters<sup>12</sup>.  $^{130}\text{Te}$  benefits both from a high natural isotopic abundance of approximately 34%<sup>13</sup> and a high  $Q_{\beta\beta}$  of 2,527.5 keV<sup>14</sup>, placing the  $0\nu\beta\beta$  decay region of interest above most natural  $\gamma$ -emitting radioactive backgrounds. The detector is an array of 988  ${}^{\text{nat}}\text{TeO}_2$  cubic crystals<sup>15</sup> (Fig. 1) of  $5 \times 5 \times 5 \text{ cm}^3$  size and  $\sim 750 \text{ g}$  mass, for a total mass of 742 kg, which corresponds to 206 kg of  $^{130}\text{Te}$ . The array is arranged as 19 towers, each comprised of 13 floors containing four crystals. The crystals are operated as cryogenic calorimeters<sup>16</sup> at a temperature of approximately 10 mK. To achieve this low-temperature environment, a novel cryogenic infrastructure—the CUORE cryostat—has been realized.

**Fig. 1: The CUORE detector.**



**a**, Rendering of the six-stage cryostat, with the pulse tubes and dilution unit, the internal low-radioactivity modern and Roman lead shields, and the array of 988 TeO<sub>2</sub> crystals (light blue). **b**, The detector after installation. The plastic ring was used during assembly for radon protection. **c**, One of the calorimeters instrumented with an NTD Ge thermistor which measures the temperature increase induced by absorbed radiation. The Si heater is used to inject pulses for thermal gain stabilization. The polytetrafluoroethylene (PTFE) supports and the gold wires instrumenting the NTD and the heater provide the thermal link between the crystal and the heat bath, that is, the Cu frames<sup>24</sup>.

In a cryogenic calorimeter, the energy deposited by impinging radiation in the absorber crystal is turned into heat, resulting in a temperature rise (Extended Data Fig. 1). Each CUORE crystal (Fig. 1c) is instrumented with a neutron-transmutation-doped germanium thermistor (NTD)<sup>17</sup> that converts thermal pulses into electric signals and a heater<sup>18</sup> to inject reference heat pulses for thermal gain stabilization<sup>19</sup>. Thermal signals can be induced by electrons emitted in  $0\nu\beta\beta$  decays but also other background radiation, for example,  $\gamma$  and  $\alpha$  particles from residual radioactive contaminants or cosmic ray muons.

CUORE is protected by several means against backgrounds that can mimic a  $0\nu\beta\beta$  decay. It is located underground at the Laboratori Nazionali del Gran Sasso (LNGS) of INFN, Italy, under a rock overburden equivalent to approximately 3,600 m of water, which shields from hadronic cosmic rays and reduces the muon flux by six orders of magnitude. Environmental  $\gamma$  backgrounds are suppressed by a 30-cm layer of low-radioactivity lead above the detector (Fig. 1), a 6-cm-thick lateral and bottom shield of <sup>210</sup>Pb-depleted ancient lead recovered from a Roman shipwreck<sup>20</sup> (Extended Data Fig. 2), and a 25-cm-thick lead shield outside the cryostat. Environmental neutrons are suppressed by a 20-cm layer of polyethylene and a thin layer of boric acid outside the external lead shield. Finally, radioactive contaminants in the crystals and in the adjacent structures are minimized by careful screening of material for radio-purity and use of high-efficiency cleaning procedures and manipulation protocols<sup>21</sup>.

## Cryogenic innovation and performance

Dilution refrigerator technology was originally proposed in the 1950s<sup>22</sup> and underwent considerable development in the 1980s driven also by the application of cryogenic calorimeters for single-particle detection<sup>23</sup>. Gradually, experimental volumes of the order of tens of litres capable of hosting cold masses of up to 60 kg at 10 mK temperature<sup>24</sup> were achieved. Ultimately, detectors were limited by the capacity, duty cycle and radio-purity of commercial or near-commercial cryogenic systems. In the context of this history, the CUORE cryostat represents a major advance in cryogenic

technology, reaching an experimental volume of approximately  $1\text{ m}^3$  and a cold mass of 1.5 t (detectors, holders, shields) at 10 mK, which corresponds to a 20-fold improvement in experimental volume and target mass compared to the previous state of the art at this temperature scale. Prior to CUORE, the ultimate temperature for comparable target masses was in the resonant-mass gravitational antenna community at 65 mK<sup>23</sup>.

The CUORE detector is hosted in a multistage cryogen-free cryostat<sup>25</sup> (Fig. 1), equipped with five pulse tube cryocoolers that avoid pre-cooling with a liquid helium bath, thus enabling a high duty cycle. A custom-designed dilution unit with a double condensing line for redundancy provides more than 4  $\mu\text{W}$  cooling power at 10 mK. The cryostat is uniquely designed to provide the necessary i) cooling power and temperature stability over a time scale of years, ii) very low radioactivity environment, and iii) extremely low-vibration conditions. As shown in Fig. 2a,b, CUORE became operational in 2017, with the initial period mostly devoted to characterization and optimization campaigns. Since 2019, the data-taking has proceeded smoothly with a duty cycle of approximately 90%. Figure 2d shows that the temperature stability achieved is at the level of 0.2% ( $\pm 3\sigma$  range) over a period in excess of one year. Such a stability is important to achieve a uniform response of all detectors over time. The CUORE calorimeters are sensitive to thermal signals and feature an intrinsic thermal fluctuation limit of approximately 0.5 keV, so any process inducing heat dissipation equal to or greater than 0.5 keV degrades the energy resolution. Mechanical vibrations can be transferred to the inner components and produce heat through friction. To minimize the impact of vibrational noise, the calorimeter array is mechanically decoupled from the cryostat by a custom suspension system. Vibrations induced by the pulse tubes at the 1.4-Hz operational frequency and its harmonics are particularly relevant. In CUORE, we actively tune the pulse tube relative phases for vibration cancellation<sup>26</sup> (Fig. 3). This solution is transferable to any cryogenic application involving signals in the same bandwidth of the pulse-tube-induced noise.

**Fig. 2: Cryogenic performance.**

---

 **figure 2**

**a**, The exposure accumulated by CUORE (teal), along with the exposure used for this analysis (orange). Parts of 2017 and 2018 were dedicated to maintenance and optimization of the cryogenic set-up. **b**, Since then, CUORE has been operating stably with a 90% duty cycle (March 2019–October 2020). **c**, Examples of temperature instabilities induced by external causes. From left to right: blackout (June 2019), earthquake in Albania of magnitude 6.4, 520 km away (November 2019), regular maintenance (July 2020), and insertion of calibration sources (September 2020). **d**, The temperature stability of CUORE over ~1 yr of continuous operation, shown by a plot of relative temperature fluctuation versus time, and a histogram of the same data. (1 t yr = 1,000 kg yr.).

**Fig. 3: Pulse tube phase optimization.**

---

 **figure 3**

**a**, Frequency spectrum of the noise for a random combination of the pulse tube phases (orange) and after the active phase tuning (teal). For reference, the frequency spectrum of physical signals is also reported. **b**, Integral of the power spectrum at the pulse tube frequency (1.4 Hz) and its harmonics before and after active noise cancellation.

CUORE now collects sensitive exposure with 984 out of 988 calorimeters, at a rate that is, to our knowledge, unprecedented for cryogenic particle detectors. Below, we describe the data treatment and  $0\nu\beta\beta$  decay search with greater than 1 t yr of  $\text{TeO}_2$  exposure.

## Data analysis and results

CUORE data are subdivided into datasets of 1–2 months of physics data, separated by a few days of calibration data collected with the detector exposed to  $^{232}\text{Th}$  and/or  $^{60}\text{Co}$  sources.

The voltage across each NTD is amplified, passed through an anti-aliasing filter, and continuously digitized with a 1-kHz sampling frequency<sup>27,28</sup>. We identify thermal pulses in the data stream and compute the pulse amplitudes, applying optimum filters that maximize the frequency-dependent signal-to-noise ratio<sup>29</sup>. To monitor and correct for possible drifts of the thermal gain of the detectors we exploit two ‘standard candles’: monoenergetic heater pulses for the calorimeters with functioning and stable heaters (>95% of the total), and events from the 2,615-keV  $^{208}\text{Tl}$  calibration line for the remainder. Drift-stabilized pulse amplitudes are converted to energy using the regularly acquired source calibration data<sup>30</sup>. We blind the  $0\nu\beta\beta$  search via a data-salting procedure that produces an artificial peak at  $Q_{\beta\beta}$ <sup>30</sup>. Once the full analysis procedure is finalized and frozen, we reverse the salting.

To simplify the analysis, we eliminate data from periods affected by high noise or failed data processing, which amounts to 5% of the exposure. Furthermore, calorimeters with greater than 19-keV full width at half maximum (FWHM) energy resolution at the 2,615-keV calibration line are discarded, adding 3% loss in exposure. In addition to these so-called base cuts, the following second-level cuts are then applied to suppress single background-like or low-quality events. Monte Carlo simulations show that approximately 88% of  $0\nu\beta\beta$  decay events release their full energy in a single crystal<sup>31</sup>. Hence, we apply an anti-coincidence cut that excludes events depositing energy in more than one crystal. Finally, we use pulse shape discrimination to eliminate pulses that are consistent with more than one energy deposit in the pulse time window, pulses with a non-physical shape, and excessively noisy pulses that survived the previous selections (Extended Data Fig. 3). The selection efficiencies are summarized in Table 1, with details provided in Methods.

### Table 1 Main parameters for the $0\nu\beta\beta$ analysis

The detector response to a monoenergetic energy deposition is an important input to the  $0\nu\beta\beta$  decay search. We empirically model the response function of each calorimeter as a sum of three equal-width Gaussians and determine the function parameters from a fit to the 2,615-keV calibration line<sup>3</sup>. As a characteristic indicator of the overall energy resolution of the calorimeters we quote the exposure-weighted harmonic mean FWHM of the detectors at this calibration line,  $7.78 \pm 0.03$  keV. All values are reported as mean  $\pm$  s.d.

We quantify the scaling of energy resolution with energy and investigate energy reconstruction bias—that is, the deviation of reconstructed  $\gamma$ -line positions from the literature values—by fitting the calorimeter response functions to prominent  $\gamma$  lines in the physics data, allowing the peak means and widths to vary in the fit. The bias is allowed to scale as a quadratic function of energy, as done in our previous result<sup>32</sup>, whereas the resolution scaling has been changed to a linear function of energy,

following studies showing that it was overparameterized by a quadratic scaling. The results, extrapolated to  $Q_{\beta\beta}$ , are an exposure-weighted harmonic mean FWHM energy resolution of  $7.8 \pm 0.5$  keV and an energy bias of less than 0.7 keV. We summarize all the relevant analysis quantities in Table 1.

Figure 4 shows the full energy spectrum along with the [2,490, 2,575] keV fit region, which contains only one background peak at 2,505.7 keV from the simultaneous absorption of two coincident  $\gamma$  rays from  $^{60}\text{Co}$  in the same crystal. We estimate that around 90% of the continuum background consists of degraded  $\alpha$  particles from radioactive contaminants of the support structure surface, and the other approximately 10% are multi-Compton scattered 2,615-keV  $\gamma$  events<sup>31,33</sup>.

**Fig. 4: Physics spectrum for 1,038.4 kg yr of  $\text{TeO}_2$  exposure.**

figure 4



We separately show the effects of the base cuts, the anti-coincidence (AC) cut, and the pulse shape discrimination (PSD). The most prominent background peaks in the spectrum are highlighted. Inset, the region of interest after all selection cuts, with the best-fit curve (solid red), the best-fit curve with the  $0\nu\beta\beta$  rate fixed to the 90% CI limit (blue), and background-only fit (black) superimposed.

We run an unbinned Bayesian fit with uniform non-negative priors on the background and  $0\nu\beta\beta$  decay rates. The fit with a background-only model—that is, excluding the  $0\nu\beta\beta$  component—yields a mean background rate of  $(1.49 \pm 0.04) \times 10^{-2}$  counts  $\text{keV}^{-1} \text{kg}^{-1} \text{yr}^{-1}$  at  $Q_{\beta\beta}$  for a corresponding median exclusion sensitivity of  $\langle T \rangle_{1/2}^{0\nu} > 2.8 \times 10^{25} \text{ years}$  (90% credibility interval (CI)). The fit with the signal-plus-background model shows no evidence for  $0\nu\beta\beta$  decay. We find the best fit at  $\Gamma_{0\nu} = (0.9 \pm 1.4) \times 10^{-26} \text{ yr}^{-1}$  and set a limit on the

process half-life of  $\langle\{T\}_\text{1/2}\rangle^0 \text{nu} > 2.2 \times 10^{25} \text{ yr}$  (90% CI). Systematic uncertainties are included as nuisance parameters and affect both the best fit and the limit by 0.8% (Extended Data Table 1). Compared to the sensitivity, the probability of getting a stronger limit is 72%. This represents, to our knowledge, the current world-leading  $0\nu\beta\beta$  sensitivity for  $^{130}\text{Te}$ , having improved in accordance with our increased exposure since our previous result<sup>32</sup>, and although the actual limit is weaker, it is well within the range of possible outcomes due to statistical fluctuations.

Under the common assumption of a light neutrino exchange mechanism, the  $^{130}\text{Te}$  half-life limit converts to a limit on the effective Majorana mass of  $m_{\beta\beta} < 90\text{--}305 \text{ meV}$ , with the spread induced by different nuclear matrix element calculations<sup>34,35,36,37,38,39,40</sup>. This limit on  $m_{\beta\beta}$  is among the strongest in the field, proving the competitiveness of the cryogenic calorimeter technique used in CUORE. CUORE will continue to take data until it reaches its designed  $^{130}\text{Te}$  exposure of 1,000 kg yr.

## Impact

We have demonstrated that the cryogenic calorimeter technique is scalable to tonne-scale detector masses and multi-year measurement campaigns, while maintaining low radioactive backgrounds. Next-generation calorimetric  $0\nu\beta\beta$  decay searches exploiting these developments are planned. Among these, CUPID (CUORE Upgrade with Particle IDentification)<sup>41</sup> will use the same cryogenic infrastructure as CUORE, replacing the  $\text{TeO}_2$  crystals with scintillating  $\{\{\text{Li}\}\}_2^{100}$   $\{\{\text{MoO}\}\}_4$  crystals and exploiting the scintillation light for greater than 100-fold active suppression of the  $\alpha$  background<sup>42,43</sup>. In parallel, the AMoRE collaboration aims to build a large-mass calorimetric  $0\nu\beta\beta$  decay experiment in Korea<sup>44</sup>. In general, the possibility to cool large detector payloads paired with the low energy thresholds achievable by cryogenic calorimeters will benefit next-generation projects at the frontier of particle physics, for example dark matter searches such as SuperCDMS<sup>45</sup> and CRESST<sup>46</sup>, and low-energy observatories exploiting CEvNS for solar and supernova neutrino studies<sup>47</sup> and neutrino flux monitoring of nuclear reactors<sup>48</sup>.

A serendipitous effect is that the cryogenic innovations pioneered by CUORE for  $0\nu\beta\beta$  decay appear to be a solution-in-waiting for the challenges faced by the relatively young, but rapidly growing, field of quantum information technology. The need to cool increasingly large arrays of qubits to less than approximately 100 mK means there is now a commercial market for large, high-cooling-power dilution refrigerators, with some featuring technological solutions derived from CUORE. Moreover, the

recent realization that ionizing radiation from natural radioactivity will be a limiting factor for the coherence time of quantum processors with an increasing number of qubits<sup>49</sup> suggests that the one-time niche, low-radioactivity ultralow-temperature cryogenics pioneered for  $0\nu\beta\beta$  decay may become mainstream in large-scale quantum computing<sup>50</sup>.

## Methods

### Optimum trigger and analysis threshold

The continuous data stream of CUORE is triggered with the optimum trigger, an algorithm based on the optimum filter<sup>51</sup> characterized by a lower threshold than a more standard derivative trigger<sup>32</sup>. A lower threshold enables us not only to reconstruct the low-energy part of the spectrum, but also yields a higher efficiency in reconstructing the events in coincidence between different calorimeters, and consequently a more precise understanding of the corresponding background components<sup>52,53</sup>.

The optimum trigger transfer function of every event is matched to the ideal signal shape, obtained as the average of good-quality pulses, so that frequency components with low signal-to-noise ratio are suppressed. A trigger is fired if the filtered signal amplitude exceeds a fixed multiple of the noise root mean square (RMS), defined separately for each calorimeter and dataset. We evaluate the energy threshold by injecting fake pulses of varying amplitude, calculated by inverting the calibration function, into the data stream. We reconstruct the stabilized amplitude of the fake pulses, fit the ratio of correctly triggered events to generated events with an error function, and use the 90% quantile as a figure of merit for the optimum trigger threshold. This approach enables monitoring of the threshold during data collection, and is based on the assumption that the signal shape is not energy dependent, that is, that the average pulse obtained from high-energy  $\gamma$  events is also a good template for events of a few keV. The distribution of energy threshold at 90% trigger efficiency is shown in Extended Data Fig. 4.

For this work we set a common analysis threshold of 40 keV, which results in >90% trigger efficiency for the majority (97%) of the calorimeters, while at the same time allowing the removal of multi-Compton events from the region of interest through the anti-coincidence cut.

### Efficiencies

The total efficiency is the product of the reconstruction, anti-coincidence, pulse shape discrimination (PSD) and containment efficiencies.

The reconstruction efficiency is the probability that a signal event is triggered, has the energy properly reconstructed, and is not rejected by the basic quality cuts requiring a stable pre-trigger voltage and only a single pulse in the signal window. It is evaluated for each calorimeter using externally flagged heater events<sup>54</sup>, which are a good approximation of signal-like events.

The anti-coincidence efficiency is the probability that a true single-crystal event correctly passes our anti-coincidence cut, instead of being wrongly vetoed owing to an accidental coincidence with an unrelated event. It is extracted as the acceptance of fully absorbed  $\gamma$  events at 1,460 keV from the electron capture decays of  $^{40}\text{K}$ , which provide a reference sample of single-crystal events.

The PSD efficiency is obtained as the average acceptance of events in the  $^{60}\text{Co}$ ,  $^{40}\text{K}$  and  $^{208}\text{Tl}$   $\gamma$  peaks that already passed the base and anti-coincidence cuts. In principle, the PSD efficiency could be different for each calorimeter, but given the limited statistics in physics data we evaluate it over all channels and over the full dataset. To account for possible variation between individual calorimeters, we compare the PSD efficiency obtained by directly summing their individual spectra with that extracted from an exposure-weighted sum of the calorimeters' spectra. We find an average  $\pm 0.3\%$  discrepancy between the two values and include it as a global systematic uncertainty in the  $0\nu\beta\beta$  fit. This takes a Gaussian prior instead of the uniform prior used in our previous result<sup>32</sup>, which had its uncertainty come from a discrepancy between two approaches that has since been resolved.

Finally, the containment efficiency is evaluated through Geant4-based Monte Carlo simulations<sup>55</sup> and accounts for the energy loss due to geometrical effects as well as bremsstrahlung.

## Principal component analysis for PSD

In this analysis we use a new algorithm based on principal component analysis (PCA) for pulse shape discrimination. The method has been developed and documented for CUPID-Mo<sup>56</sup>, and has been adapted for use in CUORE<sup>57</sup>. This technique replaces the algorithm used in previous CUORE results, which was based on six pulse shape variables<sup>30</sup>. The PCA decomposition of signal-like events pulled from  $\gamma$  calibration peaks yields a leading component similar to an average pulse, which on its own captures  $>90\%$  of the variance between pulses. We choose to treat the average pulse of each calorimeter in a dataset as if it were the leading PCA component, normalizing it like a PCA eigenvector. We can then project any event from the same channel onto

this vector and attempt to reconstruct the 10-s waveform using only this leading component. For any waveform  $\mathbf{x}$  and leading PCA component  $\mathbf{w}$  with length  $n$ , we define the reconstruction error as:

$$\text{RE} = \sqrt{\sum_{i=1}^n ((\mathbf{x})_i - (\mathbf{w})_i \cdot \mathbf{w})^2} \quad (1)$$

This reconstruction error metric measures how well an event waveform can be reconstructed using only the average pulse treated as a leading PCA component. Events that deviate from the typical expected shape of a signal waveform are poorly reconstructed and have a high reconstruction error. We normalize the reconstruction errors as a second-order polynomial function of energy on a calorimeter-dataset basis (see Extended Data Fig. 5), and cut on the normalized values by optimizing a figure of merit for signal efficiency over expected background in the  $Q_{\beta\beta}$  region of interest. Using this PCA-based method, we obtain an overall efficiency of  $(96.4 \pm 0.2)\%$  compared to the  $(94.0 \pm 0.2)\%$  from the pulse shape analysis used in our previous results, as well as a 50% reduction in the PSD systematic uncertainty from 0.6% to 0.3%.

## Statistical analysis

The high-level statistical  $0\nu\beta\beta$  decay analysis consists of an unbinned Bayesian fit on the combined data developed with the BAT software package<sup>58</sup>. The model parameters are the  $0\nu\beta\beta$  decay rate ( $\Gamma_{0\nu}$ ), a linearly sloped background, and the  $^{60}\text{Co}$  sum peak amplitude.  $\Gamma_{0\nu}$  and the  $^{60}\text{Co}$  rate are common to all datasets, with the  $^{60}\text{Co}$  rate scaled by a preset dataset-dependent factor to account for its expected decay over time. The base background rate, expressed in terms of counts  $\text{keV}^{-1} \text{kg}^{-1} \text{yr}^{-1}$ , is dataset-dependent, whereas the linear slope to the background is shared among all datasets, because any structure to the shape of the background should not vary between datasets.  $\Gamma_{0\nu}$ , the  $^{60}\text{Co}$  rate, and the background rate parameters have uniform priors that are constrained to non-negative values, whereas the linear slope to the background has a uniform prior that allows both positive and negative values.

In addition to these statistical parameters, we consider the systematic effects induced by the uncertainty on the energy bias and energy resolution<sup>59,60</sup>, the value of  $Q_{\beta\beta}$ , the natural isotopic abundance of  $^{130}\text{Te}$ , and the reconstruction, anti-coincidence, PSD and containment efficiencies. We evaluate their separate effects on the  $0\nu\beta\beta$  rate by adding nuisance parameters to the fit one at a time and studying both the effect on the posterior global mode  $\langle \hat{\Gamma}_{0\nu} \rangle$  and the marginalized 90% CI limit on  $\Gamma_{0\nu}$ .

A list of the systematics and priors is reported in Extended Data Table 1. The efficiencies and the isotopic abundance are multiplicative terms on our expected signal, so the effect of each is reported as a relative effect on  $\Gamma_{0\nu}$ . By contrast, the uncertainties on  $Q_{\beta\beta}$ , the energy bias, and the resolution scaling have a non-trivial relation to the signal rate; therefore, we report the absolute effect of each on  $\Gamma_{0\nu}$ . The dominant effect is due to the uncertainty on the energy bias and resolution scaling in physics data. We account for possible correlations between the nuisance parameters by including all of them in the fit simultaneously.

We chose a uniform prior on our physical observable of interest  $\Gamma_{0\nu}$ , which means we treat any number of signal events as equally likely. Other possible uninformative choices could be considered appropriate, as well. Because the result of any Bayesian analysis depends to some extent on the choice of the priors, we checked how other reasonable priors affect our result<sup>57</sup>. We considered: a uniform prior on  $\sqrt{\Gamma_{0\nu}}$ , equivalent to a uniform prior on  $m_{\beta\beta}$  and also equivalent to using the Jeffreys prior; a scale-invariant uniform prior on  $\log \Gamma_{0\nu}$ ; and a uniform prior on  $1/\Gamma_{0\nu}$ , equivalent to a uniform prior on  $(T_{1/2})^{0\nu}$ .

These priors are all undefined at  $\Gamma_{0\nu} = 0$ , so we impose a lower cut-off of  $\Gamma_{0\nu} > 10^{-27}$   $\text{yr}^{-1}$ , which with the given exposure corresponds to approximately one signal event. The case with a uniform prior on  $\sqrt{\Gamma_{0\nu}}$  gives the smallest effect, and strengthens the limit by 25%, whereas the flat prior on  $1/\Gamma_{0\nu}$  provides the largest effect, increasing the limit on  $(T_{1/2})^{0\nu}$  by a factor of 4. In fact, all these priors weigh the small values of  $\Gamma_{0\nu}$  more. Therefore, our choice of a flat prior on  $\Gamma_{0\nu}$  provides the most conservative result.

We compute the  $0\nu\beta\beta$  exclusion sensitivity by generating a set of  $10^4$  toy experiments with the background-model, that is, including only the  $^{60}\text{Co}$  and linear background components. The toys are split into 15 datasets with exposure and background rates obtained from the background-only fits to our actual data. We fit each toy with the signal-plus-background model, and extract the distribution of 90% CI limits, shown in Extended Data Fig. 4.

We perform the frequentist analysis using the Rolke method<sup>61</sup>, obtaining a lower limit on the process half-life of  $(T_{1/2})^{0\nu} > 2.6 \times 10^{25} \text{ yr}$  (90% CI). The profile likelihood function  $\mathcal{L}$  for  $\Gamma_{0\nu}$  is retrieved from the full Markov chain produced by the Bayesian analysis tool. The non-uniform priors on the systematic effects in the Bayesian fit are thus incorporated into the frequentist result as well. We extract a 90% confidence interval on  $\Gamma_{0\nu}$  by treating  $-2\log \mathcal{L}$  as an approximate  $\chi^2$  distribution with one degree of freedom. The lower limit on  $\Gamma_{0\nu}$

$(\{T\}_{1/2}^{1/2})^{\{0\backslash nu\}})$  comes from the corresponding upper edge of the confidence interval on  $\Gamma_{0\nu}$ . Applying the same method to the toy experiments, we find a median exclusion sensitivity of  $(\{T\}_{1/2}^{1/2})^{\{0\backslash nu\}} > 2.9 \times 10^{25} \text{ yr}$ .

## Data availability

The data generated during this analysis and shown in paper figures are available in ASCII format (CSV) as Source Data in the repository <https://cuore.lngs.infn.it/en/publications/collaborationpapers>. Additional information is available upon request by contacting the CUORE Collaboration.

## References

1. Majorana, E. Theory of the symmetry of electrons and positrons [in Italian]. *Nuovo Cim.* **14**, 171–184 (1937).
2. Fukugita, M. & Yanagida, T. Baryogenesis without grand unification. *Phys. Lett. B* **174**, 45 (1986).
3. CUORE Collaboration. . First results from CUORE: a search for lepton number violation via  $0\nu\beta\beta$  decay of  $^{130}\text{Te}$ . *Phys. Rev. Lett.* **120**, 132501 (2018).
4. Vissani, F. What is matter according to particle physics, and why try to observe its creation in a lab? *Universe* **7**, 61 (2021).
5. Bilenky, S. M. & Giunti, C. Neutrinoless double-beta decay: a probe of physics beyond the Standard Model. *Int. J. Mod. Phys. A* **30**, 1530001 (2015).
6. GERDA Collaboration. Final results of GERDA on the search for neutrinoless double- $\beta$  decay. *Phys. Rev. Lett.* **125**, 252502 (2020).
7. Majorana Collaboration. A search for neutrinoless double-beta decay in  $^{76}\text{Ge}$  with 26 kg yr of exposure from the Majorana Demonstrator. *Phys. Rev. C* **100**, 025501 (2019).
8. EXO-200 Collaboration. Search for neutrinoless double- $\beta$  decay with the complete EXO-200 dataset. *Phys. Rev. Lett.* **123**, 161802 (2019).
9. KamLAND-Zen Collaboration. Search for Majorana neutrinos near the inverted mass hierarchy region with KamLAND-Zen. *Phys. Rev. Lett.* **117**, 082503 (2016).

10. Arnaboldi, C. et al. CUORE: a Cryogenic Underground Observatory for Rare Events. *Nucl. Instrum. Meth. A* **518**, 775–798 (2004).
11. Artusa, D. R. et al. Searching for neutrinoless double-beta decay of  $^{130}\text{Te}$  with CUORE. *Adv. High Energy Phys.* **2015**, 879871 (2015).
12. Brofferio, C. & Dell’Oro, S. The saga of neutrinoless double beta decay search with  $\text{TeO}_2$  thermal detectors. *Rev. Sci. Instrum.* **89**, 121502 (2018).
13. Fehr, M. A., Rehkamper, M. & Halliday, A. N. Application of MC-ICPMS to the precise determination of tellurium isotope compositions in chondrites, iron meteorites and sulfides. *Int. J. Mass Spectrom.* **232**, 83–94 (2004).
14. Rahaman, S. et al. Double-beta decay  $Q$  values of  $^{116}\text{Cd}$  and  $^{130}\text{Te}$ . *Phys. Lett. B* **703**, 412–416 (2011).
15. Arnaboldi, C. et al. Production of high purity  $\text{TeO}_2$  single crystals for the study of neutrinoless double beta decay. *J. Cryst. Growth* **312**, 2999–3008 (2010).
16. Enss, C. & McCammon, D. Physical principles of low temperature detectors: ultimate performance limits and current detector capabilities. *J. Low Temp. Phys.* **151**, 5–24 (2008).
17. Haller, E. E., Palaio, N. P., Rodder, M., Hansen, W. L. & Kreysa, E. NTD germanium: a novel material for low temperature bolometers. In *Neutron Transmutation Doping of Semiconductor Materials* (ed. Larrabee, R. D.) 21–36 (Springer, 1984).
18. Andreotti, E. et al. Production, characterization, and selection of the heating elements for the response stabilization of the CUORE bolometers. *Nucl. Instrum. Meth. A* **664**, 161–170 (2012).
19. Alfonso, K. et al. A high precision pulse generation and stabilization system for bolometric experiments. *J. Instrum.* **13**, P02029 (2018).
20. Pattavina, L. et al. Radiopurity of an archeological Roman lead cryogenic detector. *Eur. Phys. J. A* **55**, 127 (2019).
21. Alessandria, F. et al. Validation of techniques to mitigate copper surface contamination in CUORE. *Astropart. Phys.* **45**, 13–22 (2013).
22. Proceedings of the International Conference on Low Temperature Physics. *Nature* **168**, 939–940 (1951).

23. Adams, D. Q. et al. CUORE opens the door to tonne-scale cryogenics experiments. *Prog. Part. Nucl. Phys.* **122**, 103902 (2022).
24. Alduino, C. et al. CUORE-0 detector: design, construction and operation. *J. Instrum.* **11**, P07009 (2016).
25. Alduino, C. et al. The CUORE cryostat: an infrastructure for rare event searches at millikelvin temperatures. *Cryogenics* **102**, 9–21 (2019).
26. D'Addabbo, A. et al. An active noise cancellation technique for the CUORE pulse tube cryocoolers. *Cryogenics* **93**, 56–65 (2018).
27. Arnaboldi, C. et al. A front-end electronic system for large arrays of bolometers. *J. Instrum.* **13**, P02026 (2018).
28. Di Domizio, S. et al. A data acquisition and control system for large mass bolometer arrays. *J. Instrum.* **13**, P12003 (2018).
29. Di Domizio, S., Orio, F. & Vignati, M. Lowering the energy threshold of large-mass bolometric detectors. *J. Instrum.* **6**, P02007 (2011).
30. CUORE Collaboration. Analysis techniques for the evaluation of the neutrinoless double- $\beta$  decay lifetime in  $^{130}\text{Te}$  with the CUORE-0 detector. *Phys. Rev. C* **93**, 045503 (2016).
31. Adams, D. Q. et al. Measurement of the  $2\nu\beta\beta$  decay half-life of  $^{130}\text{Te}$  with CUORE. *Phys. Rev. Lett.* **126**, 171801 (2021).
32. CUORE Collaboration. Improved limit on neutrinoless double-beta decay in  $^{130}\text{Te}$  with CUORE. *Phys. Rev. Lett.* **124**, 122501 (2020).
33. Alduino, C. et al. The projected background for the CUORE experiment. *Eur. Phys. J. C* **77**, 543 (2017).
34. Menéndez, J. et al. Disassembling the nuclear matrix elements of the neutrinoless beta beta decay. *Nucl. Phys. A* **818**, 139–151 (2009).
35. Šimkovic, F. et al.  $0\nu\beta\beta$  and  $2\nu\beta\beta$  nuclear matrix elements, quasiparticle random-phase approximation, and isospin symmetry restoration. *Phys. Rev. C* **87**, 045501 (2013).
36. López Vaquero, N., Rodríguez, T. R. & Egido, J. L. Shape and pairing fluctuations effects on neutrinoless double beta decay nuclear matrix elements. *Phys. Rev. Lett.* **111**, 142501 (2013).

37. Neacsu, A. & Horoi, M. Shell model studies of the  $^{130}\text{Te}$  neutrinoless double-beta decay. *Phys. Rev. C* **91**, 024309 (2015).
38. Yao, J. M., Song, L. S., Hagino, K., Ring, P. & Meng, J. Systematic study of nuclear matrix elements in neutrinoless double- $\beta$  decay with a beyond-mean-field covariant density functional theory. *Phys. Rev. C* **91**, 024316 (2015).
39. Barea, J., Kotila, J. & Iachello, F.  $0\nu\beta\beta$  and  $2\nu\beta\beta$  nuclear matrix elements in the interacting boson model with isospin restoration. *Phys. Rev. C* **91**, 034304 (2015).
40. Hyvärinen, J. & Suhonen, J. Nuclear matrix elements for  $0\nu\beta\beta$  decays with light or heavy Majorana-neutrino exchange. *Phys. Rev. C* **91**, 024613 (2015).
41. The CUPID Interest Group. CUPID pre-CDR. Preprint at <https://arxiv.org/abs/1907.09376> (2019).
42. Armengaud, E. et al. The CUPID-Mo experiment for neutrinoless double-beta decay: performance and prospects. *Eur. Phys. J. C* **80**, 44 (2020).
43. Azzolini, O. et al. Final result of CUPID-0 phase I in the search for the  $^{82}\text{Se}$  neutrinoless double- $\beta$  decay. *Phys. Rev. Lett.* **123**, 032501 (2019).
44. Alenkov, V. et al. First results from the AMoRE-Pilot neutrinoless double beta decay experiment. *Eur. Phys. J. C* **79**, 791 (2019).
45. SuperCDMS Collaboration. Projected sensitivity of the SuperCDMS SNOLAB experiment. *Phys. Rev. D* **95**, 082002 (2017).
46. CRESST Collaboration. First results from the CRESST-III low-mass dark matter program. *Phys. Rev. D* **100**, 102002 (2019).
47. Pattavina, L., Ferreiro Iachellini, N. & Tamborra, I. Neutrino observatory based on archaeological lead. *Phys. Rev. D* **102**, 063001 (2020).
48. NUCLEUS Collaboration. Exploring CEvNS with NUCLEUS at the Chooz nuclear power plant. *Eur. Phys. J. C* **79**, 1018 (2019).
49. Wilen, C. D. et al. Correlated charge noise and relaxation errors in superconducting qubits. *Nature* **594**, 369–373 (2021).
50. Cardani, L. et al. Reducing the impact of radioactivity on quantum circuits in a deep-underground facility. *Nat. Commun.* **12**, 2733 (2021).

51. Gatti, E. & Manfredi, P. F. Processing the signals from solid state detectors in elementary particle physics. *Riv. Nuovo Cim.* **9**, 1–146 (1986).
52. CUORE Collaboration. Low energy analysis techniques for CUORE. *Eur. Phys. J. C* **77**, 857 (2017).
53. Campani, A. et al. Lowering the energy threshold of the CUORE experiment: benefits in the surface alpha events reconstruction. *J. Low Temp. Phys.* **200**, 321–330 (2020).
54. Alessandrello, A. et al. Methods for response stabilization in bolometers for rare decays. *Nucl. Instrum. Meth. A* **412**, 454–464 (1998).
55. Alduino, C. et al. Measurement of the two-neutrino double-beta decay half-life of  $^{130}\text{Te}$  with the CUORE-0 experiment. *Eur. Phys. J. C* **77**, 13 (2017).
56. Huang, R. et al. Pulse shape discrimination in CUPID-Mo using principal component analysis. *J. Instrum.* **16**, P03032 (2021).
57. Huang, R. G. *Searching for  $0\nu\beta\beta$  Decay with CUORE and CUPID*. PhD thesis, Univ. California, Berkeley (2021).
58. Caldwell, A., Kollár, D. & Kröninger, K. BAT – the Bayesian analysis toolkit. *Comput. Phys. Commun.* **180**, 2197–2209 (2009).
59. Fantini, G. *Search for Double Beta Decay of  $^{13}\text{Te}$  to the  $0^+$  States of  $^{130}\text{Xe}$  with the CUORE Experiment*. PhD thesis, Gran Sasso Science Institute (2020).
60. Campani, A. *The Search for Neutrinoless Double Beta Decay of  $^{130}\text{Te}$  and  $^{120}\text{Te}$  with the CUORE Experiment*. PhD thesis, Univ. Genova (2021).
61. Rolke, W. A. & López, A. M. Confidence intervals and upper bounds for small signals in the presence of background noise. *Nucl. Instrum. Meth. A* **458**, 745–758 (2001).

## Acknowledgements

We thank the directors and staff of the Laboratori Nazionali del Gran Sasso and the technical staff of our laboratories. This work was supported by the Istituto Nazionale di Fisica Nucleare (INFN); the National Science Foundation under grant nos. NSF-PHY-0605119, NSF-PHY-0500337, NSF-PHY-0855314, NSF-PHY-0902171, NSF-PHY-0969852, NSF-PHY-1614611, NSF-PHY-1307204, NSF-PHY-1314881, NSF-PHY-1401832 and NSF-PHY-1913374; and Yale University. This material is also

based upon work supported by the US Department of Energy (DOE) Office of Science under contract nos. DE-AC02-05CH11231 and DE-AC52-07NA27344; by the DOE Office of Science, Office of Nuclear Physics under contract nos. DE-FG02-08ER41551, DE-FG03-00ER41138, DE-SC0012654, DE-SC0020423, DE-SC0019316; and by the EU Horizon 2020 research and innovation programme under Marie Skłodowska-Curie Grant agreement no. 754496. This research used resources of the National Energy Research Scientific Computing Center (NERSC). This work makes use of both the DIANA data analysis and APOLLO data-acquisition software packages, which were developed by the CUORICINO, CUORE, LUCIFER and CUPID-0 collaborations.

## Author information

### Author notes

1. Deceased: S. J. Freedman

### Affiliations

1. Department of Physics and Astronomy, University of South Carolina, Columbia, SC, USA

D. Q. Adams, C. Alduino, F. T. Avignone III, R. J. Creswick, C. Rosenfeld, C. Rusconi, J. Wilson & K. Wilson

2. Department of Physics and Astronomy, University of California, Los Angeles, Los Angeles, CA, USA

K. Alfonso, H. Z. Huang & L. Ma

3. INFN – Laboratori Nazionali di Legnaro, Legnaro, Italy

O. Azzolini, G. Keppel & C. Pira

4. INFN – Sezione di Bologna, Bologna, Italy

G. Bari, N. Moggi & S. Zucchelli

5. Dipartimento di Fisica, Sapienza Università di Roma, Rome, Italy

F. Bellini, G. Fantini & M. Vignati

6. INFN – Sezione di Roma, Rome, Italy

F. Bellini, L. Cardani, N. Casali, I. Dafinei, G. Fantini, F. Ferroni, S. Morganti, V. Pettinacci, C. Tomei & M. Vignati

7. INFN – Laboratori Nazionali del Gran Sasso, Assergi, Italy

G. Benato, C. Bucci, L. Canonica, L. Cappelli, E. Celi, A. D'Addabbo, V. Dompè, P. Gorla, L. Marini, C. E. Pagliarone, L. Pagnanini, L. Pattavina, S. Pirro, A. Puiu & C. Rusconi

8. Department of Physics, University of California, Berkeley, Berkeley, CA, USA

M. Beretta, S. J. Freedman, E. V. Hansen, R. G. Huang, Yu. G. Kolomensky, L. Marini, M. Sakai, V. Singh, K. J. Vetter & S. L. Wagaarachchi

9. INFN – Sezione di Milano Bicocca, Milan, Italy

M. Biassoni, A. Branca, C. Brofferio, S. Capelli, P. Carniti, D. Chiesa, M. Clemenza, O. Cremonesi, S. Dell’Oro, M. Faverzani, E. Ferri, E. Fiorini, A. Giachero, L. Gironi, C. Gotti, M. Nastasi, A. Nucciotti, I. Nutini, M. Pavan, G. Pessina, S. Pozzi, E. Previtali, M. Sisti & F. Terranova

10. Dipartimento di Fisica, Università di Milano-Bicocca, Milan, Italy

A. Branca, C. Brofferio, S. Capelli, P. Carniti, D. Chiesa, M. Clemenza, S. Dell’Oro, M. Faverzani, E. Ferri, E. Fiorini, A. Giachero, L. Gironi, M. Nastasi, A. Nucciotti, I. Nutini, M. Pavan, S. Pozzi, E. Previtali & F. Terranova

11. Center for Neutrino Physics, Virginia Polytechnic Institute and State University, Blacksburg, VA, USA

J. Camilleri, T. O’Donnell & V. Sharma

12. INFN – Sezione di Genova, Genova, Italy

A. Caminata, A. Campani, S. Copello, S. Di Domizio & M. Pallavicini

13. Dipartimento di Fisica, Università di Genova, Genova, Italy

A. Campani, S. Copello, S. Di Domizio & M. Pallavicini

14. Massachusetts Institute of Technology, Cambridge, MA, USA

L. Canonica, J. Johnston, D. Mayer, J. L. Ouellet & L. A. Winslow

15. Key Laboratory of Nuclear Physics and Ion-beam Application (MOE), Institute of Modern Physics, Fudan University, Shanghai, China

X. G. Cao, D. Q. Fang, S. H. Fu & Y. G. Ma

16. Nuclear Science Division, Lawrence Berkeley National Laboratory, Berkeley, CA, USA

S. J. Freedman, B. K. Fujikawa, Yu. G. Kolomensky, L. Marini, Y. Mei, B. Schmidt, K. J. Vetter, S. L. Wagaarachchi & B. Welliver

17. Gran Sasso Science Institute, L'Aquila, Italy

E. Celi, A. D'Addabbo, V. Dompè, F. Ferroni, L. Marini, L. Pagnanini & A. Puiu

18. INFN – Laboratori Nazionali di Frascati, Frascati, Italy

M. A. Franceschi, C. Ligi & T. Napolitano

19. IJCLab, Université Paris-Saclay, CNRS/IN2P3, Orsay, France

A. Giuliani

20. Physics Department, California Polytechnic State University, San Luis Obispo, CA, USA

T. D. Gutierrez

21. Shanghai Laboratory for Particle Physics and Cosmology, INPAC, School of Physics and Astronomy, Shanghai Jiao Tong University, Shanghai, China

K. Han

22. Wright Laboratory, Department of Physics, Yale University, New Haven, CT, USA

K. M. Heeger, R. Liu, R. H. Maruyama, J. Nikkel, S. Pagan & P. T. Surukuchi

23. Dipartimento di Fisica e Astronomia, Alma Mater Studiorum – Università di Bologna, Bologna, Italy

N. Moggi & S. Zucchelli

24. IRFU, CEA, Université Paris-Saclay, Gif-sur-Yvette, France

C. Nones

25. Lawrence Livermore National Laboratory, Livermore, CA, USA

E. B. Norman, S. Sangiorgio, N. D. Scielzo & B. S. Wang

26. Department of Nuclear Engineering, University of California, Berkeley, CA, USA

E. B. Norman & B. S. Wang

27. Dipartimento di Ingegneria Civile e Meccanica, Università degli Studi di Cassino e del Lazio Meridionale, Cassino, Italy

C. E. Pagliarone

28. Department of Physics and Astronomy, The Johns Hopkins University, Baltimore, MD, USA

D. Speller

29. INFN – Sezione di Padova, Padova, Italy

L. Taffarello

30. Engineering Division, Lawrence Berkeley National Laboratory, Berkeley, CA, USA

S. Zimmermann

## Consortia

### The CUORE Collaboration

- , D. Q. Adams
- , C. Alduino
- , K. Alfonso
- , F. T. Avignone III
- , O. Azzolini
- , G. Bari
- , F. Bellini
- , G. Benato
- , M. Beretta
- , M. Biassoni

- , A. Branca
- , C. Brofferio
- , C. Bucci
- , J. Camilleri
- , A. Caminata
- , A. Campani
- , L. Canonica
- , X. G. Cao
- , S. Capelli
- , L. Cappelli
- , L. Cardani
- , P. Carniti
- , N. Casali
- , E. Celi
- , D. Chiesa
- , M. Clemenza
- , S. Copello
- , O. Cremonesi
- , R. J. Creswick
- , A. D'Addabbo
- , I. Dafinei
- , S. Dell'Oro
- , S. Di Domizio
- , V. Dompè
- , D. Q. Fang
- , G. Fantini
- , M. Faverzani
- , E. Ferri
- , F. Ferroni
- , E. Fiorini
- , M. A. Franceschi
- , S. J. Freedman
- , S. H. Fu
- , B. K. Fujikawa
- , A. Giachero
- , L. Gironi
- , A. Giuliani
- , P. Gorla
- , C. Gotti
- , T. D. Gutierrez
- , K. Han
- , E. V. Hansen

- , K. M. Heeger
- , R. G. Huang
- , H. Z. Huang
- , J. Johnston
- , G. Keppel
- , Yu. G. Kolomensky
- , C. Ligi
- , R. Liu
- , L. Ma
- , Y. G. Ma
- , L. Marini
- , R. H. Maruyama
- , D. Mayer
- , Y. Mei
- , N. Moggi
- , S. Morganti
- , T. Napolitano
- , M. Nastasi
- , J. Nikkel
- , C. Nones
- , E. B. Norman
- , A. Nucciotti
- , I. Nutini
- , T. O'Donnell
- , J. L. Ouellet
- , S. Pagan
- , C. E. Pagliarone
- , L. Pagnanini
- , M. Pallavicini
- , L. Pattavina
- , M. Pavan
- , G. Pessina
- , V. Pettinacci
- , C. Pira
- , S. Pirro
- , S. Pozzi
- , E. Previtali
- , A. Puiu
- , C. Rosenfeld
- , C. Rusconi
- , M. Sakai
- , S. Sangiorgio

- , B. Schmidt
- , N. D. Scielzo
- , V. Sharma
- , V. Singh
- , M. Sisti
- , D. Speller
- , P. T. Surukuchi
- , L. Taffarello
- , F. Terranova
- , C. Tomei
- , K. J. Vetter
- , M. Vignati
- , S. L. Wagaarachchi
- , B. S. Wang
- , B. Welliver
- , J. Wilson
- , K. Wilson
- , L. A. Winslow
- , S. Zimmermann
- & S. Zucchelli

## Contributions

All listed authors have contributed to the present publication. The different contributions span from the design and construction of the detector and of the cryogenic system to the acquisition and analysis of data. The manuscript underwent an internal review process extended to the whole collaboration, and all authors approved its final version; the authors' names are listed alphabetically.

## Corresponding author

Correspondence to [C. Bucci](#).

## Ethics declarations

## Competing interests

The authors declare no competing interests.

## Peer review

## Peer review information

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

## Additional information

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## Extended data figures and tables

### Extended Data Fig. 1 Working principle of the cryogenic calorimeter.

Left, simplified calorimeter thermal model. The detector is modelled as a single object with heat capacity  $C$  coupled to the heat bath (with constant temperature  $T_0$ ) through the thermal conductance  $G$ . The NTD thermistor for signal readout is glued to the absorber. Right, example of a CUORE pulse from the 2,615-keV calibration line:  $T_0$  corresponds to the baseline height, the pulse amplitude is proportional to the deposited energy, and the decay time depends on the value of  $C/G$ .

### Extended Data Fig. 2 Roman lead.

Top left, the recovery of the lead bricks from the Sardinian sea. Bottom left, the ingot inscriptions were cut and preserved, and the ingot bodies were used for the CUORE internal lead shield. Right, lateral view of the internal lead shield<sup>20</sup>.

### Extended Data Fig. 3 Pulse shape discrimination.

Effect of the PSD cut on calibration data around the 2,615-keV line (left) and on physics data near  $Q_{\beta\beta}$  (right). In calibration data, the anti-coincidence is not applied in order to maximize the statistics on the  $\gamma$  peaks, and the PSD mostly removes pileup events (events with more than one energy deposit in the time window). In physics data, the PSD mostly eliminates random noise events, which can correspond to either physical events with excessive noise or to noise-induced events with non-physical pulse shapes. Such events appear randomly across the energy spectrum, so the cut mostly acts on the continuum.

### Extended Data Fig. 4 Optimum trigger and statistical analysis.

Top left, distribution of energy thresholds at 90% trigger efficiency for the optimum trigger algorithm in a single dataset. The 40-keV analysis threshold is indicated by the vertical line. Top right, 90% CI exclusion limits on  $\langle\langle T \rangle\rangle_{1/2}^{0\nu}$  from an ensemble of  $10^4$  toy experiments generated with the background-only model, with background rates obtained from the background-only fit to the data. The median exclusion sensitivity is indicated by the orange line. Bottom left, posterior probability distribution for  $\Gamma_{0\nu}$ , obtained from the Bayesian fit, with the 90% CI highlighted. Bottom right,  $\Delta\chi^2$  values obtained from the profile likelihood of  $\Gamma_{0\nu}$ , with  $\Delta\chi^2 = 0$  being the most-favoured value. The frequentist limit at 90% confidence level (CL) is indicated.

## Extended Data Fig. 5 PCA performance.

Left, example of a normalization fit of the PCA reconstruction error versus energy for a single calorimeter and dataset. The distribution contains only events that passed the other base cuts. The second-order polynomial fit is shown in orange. Right, two example pulses from this calorimeter. The actual pulse is drawn in teal, and the corresponding reconstruction obtained by the PCA is drawn in orange. The top pulse deviates from the expected shape of a good pulse and is rejected, whereas the bottom one conforms to the expected response and is accepted.

## **Extended Data Table 1 Systematics affecting the $0\nu\beta\beta$ decay analysis**

## **Supplementary information**

### Peer Review File

## **Rights and permissions**

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

### Reprints and Permissions

## About this article

### Cite this article

The CUORE Collaboration. Search for Majorana neutrinos exploiting millikelvin cryogenics with CUORE. *Nature* **604**, 53–58 (2022). <https://doi.org/10.1038/s41586-022-04497-4>

- Received: 14 April 2021
- Accepted: 01 February 2022
- Published: 06 April 2022
- Issue Date: 07 April 2022
- DOI: <https://doi.org/10.1038/s41586-022-04497-4>

### Share this article

Anyone you share the following link with will be able to read this content:

[Get shareable link](#)

Sorry, a shareable link is not currently available for this article.

[Copy to clipboard](#)

Provided by the Springer Nature SharedIt content-sharing initiative

### [Five years in the coldest fridge in the known Universe](#)

- Nick Petrić Howe
- Benjamin Thompson

Nature Podcast 06 Apr 2022

### [Cryogenic mastery aids bid to spot elusive matter creation](#)

- Jason Detwiler

News & Views 06 Apr 2022

---

This article was downloaded by **calibre** from <https://www.nature.com/articles/s41586-022-04497-4>

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

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

# Charge-density-wave-driven electronic nematicity in a kagome superconductor

- [Linpeng Nie](#) ORCID: [orcid.org/0000-0002-5495-7284](https://orcid.org/0000-0002-5495-7284)<sup>1 na1</sup>,
- [Kuanglv Sun](#) ORCID: [orcid.org/0000-0002-5017-2471](https://orcid.org/0000-0002-5017-2471)<sup>1 na1</sup>,
- [Wanru Ma](#) ORCID: [orcid.org/0000-0002-8390-2119](https://orcid.org/0000-0002-8390-2119)<sup>2 na1</sup>,
- [Dianwu Song](#)<sup>1</sup>,
- [Lixuan Zheng](#)<sup>1</sup>,
- [Zuowei Liang](#)<sup>1</sup>,
- [Ping Wu](#)<sup>2</sup>,
- [Fanghang Yu](#)<sup>1</sup>,
- [Jian Li](#)<sup>1</sup>,
- [Min Shan](#)<sup>1</sup>,
- [Dan Zhao](#)<sup>1</sup>,
- [Shunjiao Li](#)<sup>1</sup>,
- [Baolei Kang](#)<sup>1</sup>,
- [Zhimian Wu](#)<sup>1</sup>,
- [Yanbing Zhou](#)<sup>1</sup>,
- [Kai Liu](#)<sup>1</sup>,
- [Ziji Xiang](#) ORCID: [orcid.org/0000-0001-7132-6591](https://orcid.org/0000-0001-7132-6591)<sup>1</sup>,
- [Jianjun Ying](#)<sup>2</sup>,
- [Zhenyu Wang](#) ORCID: [orcid.org/0000-0002-4857-7234](https://orcid.org/0000-0002-4857-7234)<sup>2,3</sup>,
- [Tao Wu](#) ORCID: [orcid.org/0000-0001-9805-4434](https://orcid.org/0000-0001-9805-4434)<sup>1,2,3,4</sup> &
- [Xianhui Chen](#) ORCID: [orcid.org/0000-0001-6947-1407](https://orcid.org/0000-0001-6947-1407)<sup>1,2,3,4,5</sup>

- 8868 Accesses
- 2 Citations
- 79 Altmetric
- [Metrics details](#)

## Subjects

- [Electronic properties and materials](#)
- [Superconducting properties and materials](#)

## Abstract

Electronic nematicity, in which rotational symmetry is spontaneously broken by electronic degrees of freedom, has been demonstrated as a ubiquitous phenomenon in correlated quantum fluids including high-temperature superconductors and quantum Hall systems<sup>1,2</sup>. Notably, the electronic nematicity in high-temperature superconductors exhibits an intriguing entanglement with superconductivity, generating complicated superconducting pairing and intertwined electronic orders. Recently, an unusual competition between superconductivity and a charge-density-wave (CDW) order has been found in the  $AV_3Sb_5$  ( $A = K, Rb, Cs$ ) family with two-dimensional vanadium kagome nets<sup>3,4,5,6,7,8</sup>. Whether these phenomena involve electronic nematicity is still unknown. Here we report evidence for the existence of electronic nematicity in  $CsV_3Sb_5$ , using a combination of elastoresistance measurements, nuclear magnetic resonance (NMR) and scanning tunnelling microscopy/spectroscopy (STM/S). The temperature-dependent elastoresistance coefficient ( $m_{11}$  minus  $m_{12}$ ) and NMR spectra demonstrate that, besides a  $C_2$  structural distortion of the  $2a_0 \times 2a_0$  supercell owing to out-of-plane modulation, considerable nematic fluctuations emerge immediately below the CDW transition (approximately 94 kelvin) and finally a nematic transition occurs below about 35 kelvin. The STM experiment directly visualizes the  $C_2$ -structure-pinned long-range nematic order below the nematic transition temperature, suggesting a novel

nematicity described by a three-state Potts model. Our findings indicate an intrinsic electronic nematicity in the normal state of  $\text{CsV}_3\text{Sb}_5$ , which sets a new paradigm for revealing the role of electronic nematicity on pairing mechanism in unconventional superconductors.

This is a preview of subscription content

## Access options

Subscribe to Nature+

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

\$29.99

monthly

[Subscribe](#)

Subscribe to Journal

Get full journal access for 1 year

\$199.00

only \$3.90 per issue

[Subscribe](#)

All prices are NET prices.

VAT will be added later in the checkout.

Tax calculation will be finalised during checkout.

Buy article

Get time limited or full article access on ReadCube.

\$32.00

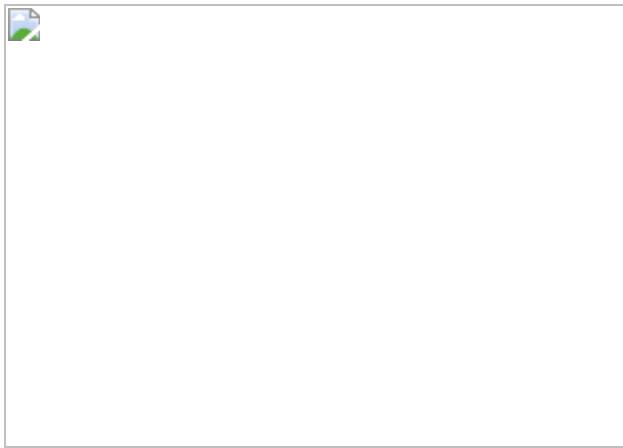
[Buy](#)

All prices are NET prices.

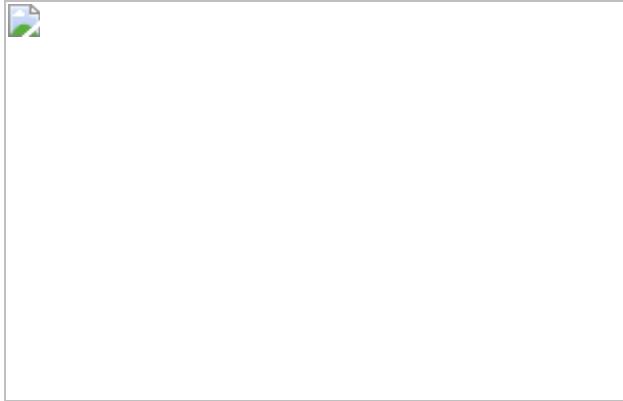
### **Additional access options:**

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

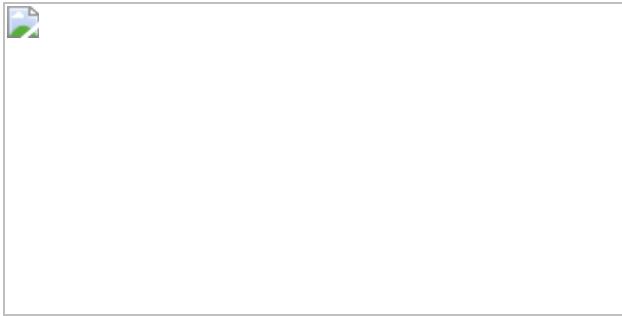
**Fig. 1: Visualizing the intra-unit-cell rotational symmetry breaking of the  $2a_0 \times 2a_0$  CDW state.**



**Fig. 2:  $^{51}\text{V}$  NMR evidence for  $C_6$  rotational symmetry breaking and electronic nematicity.**



**Fig. 3: Evidence for CDW-driven electronic nematicity from elastoresistance measurement.**



## Code availability

The code used for STM data analysis is available from the corresponding author upon reasonable request.

## Data availability

The data supporting the findings of this study are available from the corresponding author upon reasonable request. [Source data](#) are provided with this paper.

## References

1. Fradkin, E., Kivelson, S., Lawler, M. J., Eisenstein, J. P. & Mackenzie, A. P. Nematic Fermi fluids in condensed matter physics. *Annu. Rev. Condens. Matter Phys.* **1**, 153–178 (2010).
2. Fernandes, R. M., Orth, P. P. & Schmalian, J. Intertwined vestigial order in quantum materials: nematicity and beyond. *Annu. Rev. Condens. Matter Phys.* **10**, 133–154 (2019).
3. Ortiz, B. R. et al. New kagome prototype materials: discovery of  $\text{KV}_3\text{Sb}_5$ ,  $\text{RbV}_3\text{Sb}_5$ , and  $\text{CsV}_3\text{Sb}_5$ . *Phys. Rev. Mater.* **3**, 094407 (2019).
4. Ortiz, B. R. et al.  $\text{CsV}_3\text{Sb}_5$ : a  $\mathbb{Z}_2$  topological kagome metal with a superconducting ground state. *Phys. Rev. Lett.* **125**, 247002 (2020).

5. Chen, K. Y. et al. Double superconducting dome and triple enhancement of  $T_c$  in the kagome superconductor  $\text{CsV}_3\text{Sb}_5$  under high pressure. *Phys. Rev. Lett.* **126**, 247001 (2021).
6. Yu, F. H. et al. Unusual competition of superconductivity and charge-density-wave state in a compressed topological kagome metal. *Nat. Commun.* **12**, 3645 (2021).
7. Zhao, H. et al. Cascade of correlated electron states in the kagome superconductor  $\text{CsV}_3\text{Sb}_5$ . *Nature* **599**, 216–221 (2021).
8. Chen, H. et al. Roton pair density wave in a strong-coupling kagome superconductor. *Nature* **599**, 222–228 (2021).
9. Martin, I. & Batista, C. D. Itinerant electron-driven chiral magnetic ordering and spontaneous quantum Hall effect in triangular lattice models. *Phys. Rev. Lett.* **101**, 156402 (2008).
10. Ko, W.-H., Lee, P. A. & Wen, X.-G. Doped kagome system as exotic superconductor. *Phys. Rev. B* **79**, 214502 (2009).
11. Guo, H.-M. & Franz, M. Topological insulator on the kagome lattice. *Phys. Rev. B* **80**, 113102 (2009).
12. Nandkishore, R., Levitov, L. S. & Chubukov, A. V. Chiral superconductivity from repulsive interactions in doped graphene. *Nat. Phys.* **8**, 158–163 (2012).
13. Yu, S.-L. & Li, J.-X. Chiral superconducting phase and chiral spin-density-wave phase in a Hubbard model on the kagome lattice. *Phys. Rev. B* **85**, 144402 (2012).
14. Kiesel, M. L., Platt, C. & Thomale, R. Unconventional Fermi surface instabilities in the kagome Hubbard model. *Phys. Rev. Lett.* **110**, 126405 (2013).
15. Wang, W.-S., Li, Z.-Z., Xiang, Y.-Y. & Wang, Q.-H. Competing electronic orders on kagome lattices at van Hove filling. *Phys. Rev. B*

**87**, 115135 (2013).

16. Kang, M. et al. Twofold van Hove singularity and origin of charge order in topological kagome superconductor  $\text{CsV}_3\text{Sb}_5$ . *Nat. Phys.* <https://doi.org/10.1038/s41567-021-01451-5> (2022).
17. Hu, Y. et al. Rich nature of Van Hove singularities in kagome superconductor  $\text{CsV}_3\text{Sb}_5$ . Preprint at <https://arxiv.org/abs/2106.05922> (2021).
18. Jiang, Y.-X. et al. Unconventional chiral charge order in kagome superconductor  $\text{KV}_3\text{Sb}_5$ . *Nat. Mater.* **20**, 1353–1357 (2021).
19. Liang, Z. et al. Three-dimensional charge density wave and surface-dependent vortex-core states in a kagome superconductor  $\text{CsV}_3\text{Sb}_5$ . *Phys. Rev. X* **11**, 031026 (2021).
20. Li, H. et al. Observation of unconventional charge density wave without acoustic phonon anomaly in kagome superconductors  $AV_3\text{Sb}_5$  ( $A = \text{Rb}, \text{Cs}$ ). *Phys. Rev. X* **11**, 031050 (2021).
21. Tan, H., Liu, Y., Wang, Z. & Yan, B. Charge density waves and electronic properties of superconducting kagome metals. *Phys. Rev. Lett.* **127**, 046401 (2021).
22. Song, D. et al. Orbital ordering and fluctuations in a kagome superconductor  $\text{CsV}_3\text{Sb}_5$ . *Sci. China-Phys. Mech. Astron.* **65**, 247462 (2022).
23. Feng, X., Jiang, K., Wang, Z. & Hu, J. Chiral flux phase in the kagome superconductor  $AV_3\text{Sb}_5$ . *Sci. Bull.* **66**, 1384–1388 (2021).
24. Mielke, C. III et al. Time-reversal symmetry-breaking charge order in a correlated kagome superconductor. *Nature* **602**, 245–250 (2022).
25. Yu, L. et al. Evidence of a hidden flux phase in the topological kagome metal  $\text{CsV}_3\text{Sb}_5$ . Preprint at <https://arxiv.org/abs/2107.10714> (2021).

26. Yang, S.-Y. et al. Giant, unconventional anomalous Hall effect in the metallic frustrated magnet candidate,  $\text{KV}_3\text{Sb}_5$ . *Sci. Adv.* **6**, eabb6003 (2020).
27. Yu, F. H. et al. Concurrence of anomalous Hall effect and charge density wave in a superconducting topological kagome metal. *Phys. Rev. B* **104**, L041103 (2021).
28. Li, H. et al. Rotation symmetry breaking in the normal state of a kagome superconductor  $\text{KV}_3\text{Sb}_5$ . *Nat. Phys.* <https://doi.org/10.1038/s41567-021-01479-7> (2022).
29. Xiang, Y. et al. Twofold symmetry of  $c$ -axis resistivity in topological kagome superconductor  $\text{CsV}_3\text{Sb}_5$  with in-plane rotating magnetic field. *Nat. Commun.* **12**, 6727 (2021).
30. Park, T., Ye, M. & Balents, L. Electronic instabilities of kagome metals: saddle points and Landau theory. *Phys. Rev. B* **104**, 035142 (2021).
31. Christensen, M. H., Birol, T., Andersen, B. M. & Fernandes, R. M. Theory of the charge density wave in  $AV_3\text{Sb}_5$  kagome metals. *Phys. Rev. B* **104**, 214513 (2021).
32. Ortiz, B. R. et al. Fermi surface mapping and the nature of charge-density-wave order in the kagome superconductor  $\text{CsV}_3\text{Sb}_5$ . *Phys. Rev. X* **11**, 041030 (2021).
33. Chu, J.-H., Kuo, H.-H., Analytis, J. G. & Fisher, I. R. Divergent nematic susceptibility in an iron arsenide superconductor. *Science* **337**, 710–712 (2012).
34. Kuo, H.-H., Chu, J.-H., Palmstrom, J. C., Kivelson, S. A. & Fisher, I. R. Ubiquitous signatures of nematic quantum criticality in optimally doped Fe-based superconductors. *Science* **352**, 958–962 (2016).

35. Hinkov, V. et al. Electronic liquid crystal state in the high-temperature superconductor  $\text{YBa}_2\text{Cu}_3\text{O}_{6.45}$ . *Science* **319**, 597–600 (2008).
36. Daou, R. et al. Broken rotational symmetry in the pseudogap phase of a high- $T_c$  superconductor. *Nature* **463**, 519–522 (2010).
37. Sato, Y. et al. Thermodynamic evidence for a nematic phase transition at the onset of the pseudogap in  $\text{YBa}_2\text{Cu}_3\text{O}_y$ . *Nat. Phys.* **13**, 1074–1078 (2017).
38. Agterberg, D. F. et al. The physics of pair-density waves: cuprate superconductors and beyond. *Annu. Rev. Condens. Matter Phys.* **11**, 231–270 (2020).
39. Fernandes, R. M. & Venderbos, J. W. F. Nematicity with a twist: rotational symmetry breaking in a moiré superlattice. *Sci. Adv.* **6**, eaba8834 (2020).
40. Zhao, C. C. et al. Nodal superconductivity and superconducting domes in the topological kagome metal  $\text{CsV}_3\text{Sb}_5$ . Preprint at <https://arxiv.org/abs/2102.08356> (2021).
41. Xu, H.-S. et al. Multiband superconductivity with sign-preserving order parameter in kagome superconductor  $\text{CsV}_3\text{Sb}_5$ . *Phys. Rev. Lett.* **127**, 187004 (2021).
42. Mu, C. et al. S-wave superconductivity in kagome metal  $\text{CsV}_3\text{Sb}_5$  revealed by  $^{121}/^{123}\text{Sb}$  NQR and  $^{51}\text{V}$  NMR measurements. *Chin. Phys. Lett.* **38**, 077402 (2021).
43. Duan, W. et al. Nodeless superconductivity in the kagome metal  $\text{CsV}_3\text{Sb}_5$ . *Sci. China Phys. Mech. Astron.* **64**, 107462 (2021).
44. Fernandes, R. M. & Millis, A. J. Nematicity as a probe of superconducting pairing in iron-based superconductors. *Phys. Rev. Lett.* **111**, 127001 (2013).

45. Tao, W. et al. Magnetic-field-induced charge-stripe order in the high-temperature superconductor  $\text{YBa}_2\text{Cu}_3\text{O}_y$ . *Nature* **477**, 191–194 (2011).
46. Berthier, C., Jérôme, D. & Molinie, P. NMR study on a 2H-NbSe<sub>2</sub> single crystal: a microscopic investigation of the charge density waves state. *J. Phys. C* **11**, 797–814 (1978).
47. Hosoi, S. et al. Nematic quantum critical point without magnetism in  $\text{FeSe}_{1-x}\text{S}_x$  superconductors. *Proc. Natl. Acad. Sci. USA* **113**, 8139–8143 (2016).
48. Ishida, K. et al. Divergent nematic susceptibility near the pseudogap critical point in a cuprate superconductor. *J. Phys. Soc. Jpn.* **89**, 064707 (2020).
49. Eckberg, C. et al. Sixfold enhancement of superconductivity in a tunable electronic nematic system. *Nat. Phys.* **16**, 346–350 (2020).
50. Shapiro, M. C., Hlobil, P., Hristov, A. T., Maharaj, A. V. & Fisher, I. R. Symmetry constraints on the elastoresistivity tensor. *Phys. Rev. B* **92**, 235147 (2015).
51. Lawler, M. J. et al. Intra-unit-cell electronic nematicity of the high- $T_c$  copper-oxide pseudogap states. *Nature* **466**, 347–351 (2010).
52. Li, H. et al. Spatial symmetry constraint of charge-ordered kagome superconductor  $\text{CsV}_3\text{Sb}_5$ . Preprint at <https://arxiv.org/abs/2109.03418> (2021).

## Acknowledgements

We thank K. Jiang, J. He, L. Jiao, X. Liu and V. Madhavan for discussions. We thank J. Jiang for assistance with Laue diffraction measurement. This work is supported by the National Key R&D Program of the MOST of China (grant no. 2017YFA0303000), the National Natural Science Foundation of China (grant nos 11888101, 12034004, 12074364), the Strategic Priority Research Program of the Chinese Academy of Sciences

(grant no. XDB25000000), the Anhui Initiative in Quantum Information Technologies (grant no. AHY160000), the Collaborative Innovation Program of Hefei Science Center, CAS (grant no. 2019HSC-CIP007).

## Author information

### Author notes

1. These authors contributed equally: Linpeng Nie, Kuanglv Sun, Wanru Ma

### Affiliations

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

Linpeng Nie, Kuanglv Sun, Dianwu Song, Lixuan Zheng, Zuowei Liang, Fanghang Yu, Jian Li, Min Shan, Dan Zhao, Shunjiao Li, Baolei Kang, Zhimian Wu, Yanbing Zhou, Kai Liu, Ziji Xiang, Tao Wu & Xianhui Chen

2. CAS Key Laboratory of Strongly-coupled Quantum Matter Physics, Department of Physics, University of Science and Technology of China, Hefei, China

Wanru Ma, Ping Wu, Jianjun Ying, Zhenyu Wang, Tao Wu & Xianhui Chen

3. CAS Center for Excellence in Superconducting Electronics (CENSE), Shanghai, China

Zhenyu Wang, Tao Wu & Xianhui Chen

4. Collaborative Innovation Center of Advanced Microstructures, Nanjing University, Nanjing, China

Tao Wu & Xianhui Chen

5. CAS Center for Excellence in Quantum Information and Quantum Physics, Hefei, China

Xianhui Chen

## Contributions

T.W., Z. Wang and X.C. conceived the experiments. W.M., P.W., Z.L. and Z. Wang performed STM experiments. L.N., D.S., L.Z., J.L., M.S., D.Z., S.L., B.K., Z. Wu, Y.Z., K.L. and T.W. performed NMR measurements. K.S., L.N. and T.W. performed elastoresistance measurements. F.Y. and J.Y. grew the single crystals. L.N., K.S., Z.X., Z. Wang, T.W. and X.C. interpreted the results and wrote the manuscript. All authors discussed the results and commented on the manuscript.

## Corresponding authors

Correspondence to [Zhenyu Wang](#), [Tao Wu](#) or [Xianhui Chen](#).

## Ethics declarations

## Competing interests

The authors declare no competing interests.

## Peer review

## Peer review information

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

## Additional information

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## Supplementary information

### Supplementary Information

This file contains Supplementary Sections 1–16, including 18 Supplementary Figures, Supplementary Tables 1–2 and Supplementary References; see contents page for details.

## Source data

### Source Data Fig. 1

### Source Data Fig. 2

### Source Data Fig. 3

## Rights and permissions

### Reprints and Permissions

## About this article

### Cite this article

Nie, L., Sun, K., Ma, W. *et al.* Charge-density-wave-driven electronic nematicity in a kagome superconductor. *Nature* **604**, 59–64 (2022).  
<https://doi.org/10.1038/s41586-022-04493-8>

- Received: 02 September 2021
- Accepted: 26 January 2022

- Published: 09 February 2022
- Issue Date: 07 April 2022
- DOI: <https://doi.org/10.1038/s41586-022-04493-8>

## Share this article

Anyone you share the following link with will be able to read this content:

[Get shareable link](#)

Sorry, a shareable link is not currently available for this article.

[Copy to clipboard](#)

Provided by the Springer Nature SharedIt content-sharing initiative

## [\*\*Electrons go loopy in a family of superconductors\*\*](#)

- Morten H. Christensen
- Turan Birol

News & Views 09 Feb 2022

---

This article was downloaded by **calibre** from <https://www.nature.com/articles/s41586-022-04493-8>

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

- Article
- [Published: 06 April 2022](#)

# Ultrathin ferroic HfO<sub>2</sub>–ZrO<sub>2</sub> superlattice gate stack for advanced transistors

- [Suraj S. Cheema](#)<sup>1</sup><sup>na1</sup>,
- [Nirmaan Shanker](#) ORCID: [orcid.org/0000-0002-4102-9665](https://orcid.org/0000-0002-4102-9665)<sup>2</sup><sup>na1</sup>,
- [Li-Chen Wang](#) ORCID: [orcid.org/0000-0003-0703-5030](https://orcid.org/0000-0003-0703-5030)<sup>1</sup>,
- [Cheng-Hsiang Hsu](#) ORCID: [orcid.org/0000-0002-6863-5606](https://orcid.org/0000-0002-6863-5606)<sup>2</sup>,
- [Shang-Lin Hsu](#)<sup>2</sup>,
- [Yu-Hung Liao](#)<sup>2</sup>,
- [Matthew San Jose](#)<sup>3</sup>,
- [Jorge Gomez](#)<sup>3</sup>,
- [Wriddhi Chakraborty](#)<sup>3</sup>,
- [Wenshen Li](#)<sup>2</sup>,
- [Jong-Ho Bae](#) ORCID: [orcid.org/0000-0002-1786-7132](https://orcid.org/0000-0002-1786-7132)<sup>2</sup>,
- [Steve K. Volkman](#)<sup>4</sup>,
- [Daewoong Kwon](#)<sup>2</sup>,
- [Yoonsoo Rho](#)<sup>5</sup>,
- [Gianni Pinelli](#)<sup>6</sup>,
- [Ravi Rastogi](#)<sup>6</sup>,
- [Dominick Pipitone](#)<sup>6</sup>,
- [Corey Stull](#)<sup>6</sup>,
- [Matthew Cook](#)<sup>6</sup>,
- [Brian Tyrrell](#)<sup>6</sup>,
- [Vladimir A. Stoica](#)<sup>7</sup>,
- [Zhan Zhang](#)<sup>8</sup>,

- [John W. Freeland](#)<sup>8</sup>,
- [Christopher J. Tassone](#)<sup>9</sup>,
- [Apurva Mehta](#) ORCID: orcid.org/0000-0003-0870-6932<sup>9</sup>,
- [Ghazal Saheli](#)<sup>10</sup>,
- [David Thompson](#)<sup>10</sup>,
- [Dong Ik Suh](#)<sup>11</sup>,
- [Won-Tae Koo](#)<sup>11</sup>,
- [Kab-Jin Nam](#)<sup>12</sup>,
- [Dong Jin Jung](#)<sup>12</sup>,
- [Woo-Bin Song](#)<sup>12</sup>,
- [Chung-Hsun Lin](#)<sup>13</sup>,
- [Seunggeol Nam](#)<sup>14</sup>,
- [Jinseong Heo](#) ORCID: orcid.org/0000-0003-2530-488X<sup>14</sup>,
- [Narendra Parihar](#)<sup>15</sup>,
- [Costas P. Grigoropoulos](#) ORCID: orcid.org/0000-0002-8505-4037<sup>5</sup>,
- [Padraic Shafer](#) ORCID: orcid.org/0000-0001-9363-2557<sup>16</sup>,
- [Patrick Fay](#)<sup>3</sup>,
- [Ramamoorthy Ramesh](#) ORCID: orcid.org/0000-0003-0524-1332<sup>1,17,18</sup>,
- [Souvik Mahapatra](#)<sup>15</sup>,
- [Jim Ciston](#) ORCID: orcid.org/0000-0002-8774-5747<sup>19</sup>,
- [Suman Datta](#)<sup>3</sup>,
- [Mohamed Mohamed](#)<sup>6</sup>,
- [Chenming Hu](#)<sup>2</sup> &
- [Sayeef Salahuddin](#) <sup>2,18</sup>

*Nature* volume **604**, pages 65–71 (2022)

- 2158 Accesses
- 117 Altmetric
- [Metrics details](#)

## Subjects

- [Electrical and electronic engineering](#)
- [Electronic devices](#)
- [Electronic properties and materials](#)

## Abstract

With the scaling of lateral dimensions in advanced transistors, an increased gate capacitance is desirable both to retain the control of the gate electrode over the channel and to reduce the operating voltage<sup>1</sup>. This led to a fundamental change in the gate stack in 2008, the incorporation of high-dielectric-constant HfO<sub>2</sub> (ref. <sup>2</sup>), which remains the material of choice to date. Here we report HfO<sub>2</sub>–ZrO<sub>2</sub> superlattice heterostructures as a gate stack, stabilized with mixed ferroelectric–antiferroelectric order, directly integrated onto Si transistors, and scaled down to approximately 20 ångströms, the same gate oxide thickness required for high-performance transistors. The overall equivalent oxide thickness in metal–oxide–semiconductor capacitors is equivalent to an effective SiO<sub>2</sub> thickness of approximately 6.5 ångströms. Such a low effective oxide thickness and the resulting large capacitance cannot be achieved in conventional HfO<sub>2</sub>-based high-dielectric-constant gate stacks without scavenging the interfacial SiO<sub>2</sub>, which has adverse effects on the electron transport and gate leakage current<sup>3</sup>. Accordingly, our gate stacks, which do not require such scavenging, provide substantially lower leakage current and no mobility degradation. This work demonstrates that ultrathin ferroic HfO<sub>2</sub>–ZrO<sub>2</sub> multilayers, stabilized with competing ferroelectric–antiferroelectric order in the two-nanometre-thickness regime, provide a path towards advanced gate oxide stacks in electronic devices beyond conventional HfO<sub>2</sub>-based high-dielectric-constant materials.

This is a preview of subscription content

## Access options

Subscribe to Nature+

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

\$29.99

monthly

[Subscribe](#)

Subscribe to Journal

Get full journal access for 1 year

\$199.00

only \$3.90 per issue

[Subscribe](#)

All prices are NET prices.

VAT will be added later in the checkout.

Tax calculation will be finalised during checkout.

Buy article

Get time limited or full article access on ReadCube.

\$32.00

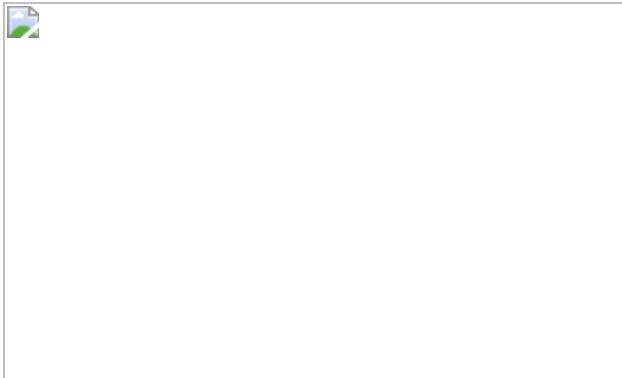
[Buy](#)

All prices are NET prices.

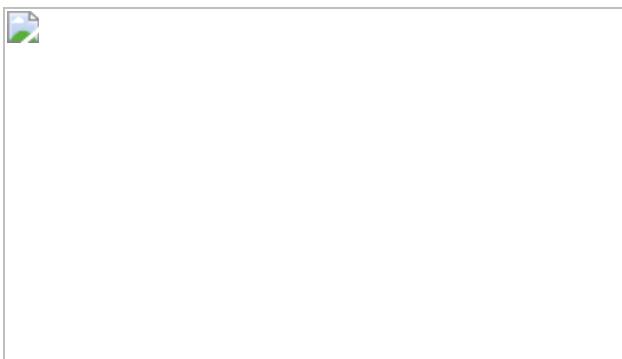
### **Additional access options:**

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

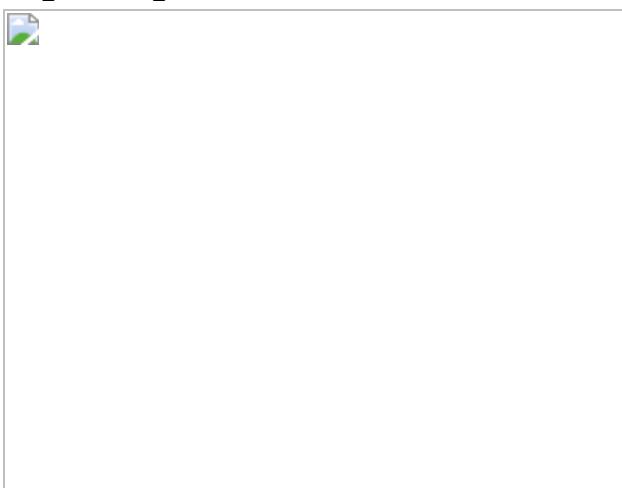
**Fig. 1: Atomic-scale design of negative capacitance in ultrathin HfO<sub>2</sub>–ZrO<sub>2</sub>.**



**Fig. 2: Enhanced capacitance in ultrathin HfO<sub>2</sub>–ZrO<sub>2</sub> mixed-ferroic heterostructures.**



**Fig. 3: Device performance benefits from using ultrathin mixed-ferroic HfO<sub>2</sub>–ZrO<sub>2</sub> gate stacks.**



## Data availability

The experimental data contained in the manuscript are available for download at <https://doi.org/10.5281/zenodo.5797030>.

## References

1. Theis, T. N. & Wong, H.-S. P. The end of Moore's law: a new beginning for information technology. *Comput. Sci. Eng.* **19**, 41–50 (2017).
2. Schlom, D. G., Guha, S. & Datta, S. Gate oxides beyond SiO<sub>2</sub>. *MRS Bull.* **33**, 1017–1025 (2008).
3. Ando, T. Ultimate scaling of high- $\kappa$  gate dielectrics: higher- $\kappa$  or interfacial layer scavenging? *Materials* **5**, 478–500 (2012).
4. Salahuddin, S., Ni, K. & Datta, S. The era of hyper-scaling in electronics. *Nat. Electron.* **1**, 442–450 (2018).
5. Shulaker, M. M. et al. Three-dimensional integration of nanotechnologies for computing and data storage on a single chip. *Nature* **547**, 74–78 (2017).
6. Wong, H.-S. & Salahuddin, S. Memory leads the way to better computing. *Nat. Nanotechnol.* **10**, 191–194 (2015).
7. Del Alamo, J. A. Nanometre-scale electronics with III–V compound semiconductors. *Nature* **479**, 317–323 (2011).
8. Butler, S. Z. et al. Progress, challenges, and opportunities in two-dimensional materials beyond graphene. *ACS Nano* **7**, 2898–2926 (2013).
9. Khan, A. I., Keshavarzi, A. & Datta, S. The future of ferroelectric field-effect transistor technology. *Nat. Electron.* **3**, 588–597 (2020).
10. Dutta, S. et al. Monolithic 3D integration of high endurance multi-bit ferroelectric FET for accelerating compute-in-memory. In *2020 IEEE International Electron Devices Meeting (IEDM)* 36.4.1–36.4.4 (IEEE, 2020).

11. Salahuddin, S. & Datta, S. Use of negative capacitance to provide voltage amplification for low power nanoscale devices. *Nano Lett.* **8**, 405–410 (2008).
12. Böscke, T. S., Müller, J., Bräuhaus, D., Schröder, U. & Böttger, U. Ferroelectricity in hafnium oxide thin films. *Appl. Phys. Lett.* **99**, 102903 (2011).
13. Cheema, S. S. et al. Enhanced ferroelectricity in ultrathin films grown directly on silicon. *Nature* **580**, 478–482 (2020).
14. Lee, H.-J. et al. Scale-free ferroelectricity induced by flat phonon bands in  $\text{HfO}_2$ . *Science* **369**, 1343–1347 (2020).
15. Noheda, B. & Iniguez, J. A key piece of the ferroelectric hafnia puzzle. *Science* **369**, 1300–1301 (2020).
16. Ando, T. et al. Understanding mobility mechanisms in extremely scaled  $\text{HfO}_2$  (EOT 0.42 nm) using remote interfacial layer scavenging technique and  $V_t$ -tuning dipoles with gatefirst process. In *2009 IEEE International Electron Devices Meeting (IEDM)* 17.1 (IEEE, 2009).
17. Wong, H. & Iwai, H. On the scaling of subnanometer EOT gate dielectrics for ultimate nano CMOS technology. *Microelectron. Eng.* **138**, 57–76 (2015).
18. Narasimha, S. et al. 22 nm high-performance SOI technology featuring dual-embedded stressors, Epi-Plate High-K deep-trench embedded DRAM and self-aligned via 15LM BEOL. In *2012 International Electron Devices Meeting* 3.3.1–3.3.4 (IEEE, 2012).
19. Huang, J. Gate first high-k/metal gate stacks with zero  $\text{SiO}_x$  interface achieving EOT=0.59 nm for 16 nm application. In *2009 Symposium on VLSI Technology* 34–35 (IEEE, 2009).
20. Yeo, Y.-C., King, T.-J. & Hu, C. Direct tunneling leakage current and scalability of alternative gate dielectrics. *Appl. Phys. Lett.* **81**, 2091–2093 (2002).

21. Kittel, C. Theory of antiferroelectric crystals. *Phys. Rev.* **82**, 729–732 (1951).
22. Materlik, R., Künneth, C. & Kersch, A. The origin of ferroelectricity in  $\text{Hf}_{1-x}\text{Zr}_x\text{O}_2$ : a computational investigation and a surface energy model. *J. Appl. Phys.* **117**, 134109 (2015).
23. Reyes-Lillo, S. E., Garrity, K. F. & Rabe, K. M. Antiferroelectricity in thin-film  $\text{ZrO}_2$  from first principles. *Phys. Rev. B* **90**, 140103 (2014).
24. Qi, Y. & Rabe, K. M. Phase competition in  $\text{HfO}_2$  with applied electric field from first principles. *Phys. Rev. B* **102**, 214108 (2020).
25. Lomenzo, P. D., Richter, C., Mikolajick, T. & Schroeder, U. Depolarization as driving force in antiferroelectric hafnia and ferroelectric wake-up. *ACS Appl. Electron. Mater.* **2**, 1583–1595 (2020).
26. Hoffmann, M. et al. Unveiling the double-well energy landscape in a ferroelectric layer. *Nature* **565**, 464–467 (2019).
27. Íñiguez, J., Zubko, P., Luk'yanchuk, I. & Cano, A. Ferroelectric negative capacitance. *Nat. Rev. Mater.* **4**, 243–256 (2019).
28. Li, F., Zhang, S., Damjanovic, D., Chen, L.-Q. & Shrout, T. R. Local structural heterogeneity and electromechanical responses of ferroelectrics: learning from relaxor ferroelectrics. *Adv. Funct. Mater.* **28**, 1801504 (2018).
29. Khan, A. et al. Experimental evidence of ferroelectric negative capacitance in nanoscale heterostructures. *Appl. Phys. Lett.* **99**, 113501 (2011).
30. Yadav, A. K. et al. Spatially resolved steady-state negative capacitance. *Nature* **565**, 468–471 (2019).
31. Das, S. et al. Local negative permittivity and topological phase transition in polar skyrmions. *Nat. Mater.* **20**, 194–201 (2021).

32. Müller, J. et al. Ferroelectricity in simple binary  $ZrO_2$  and  $HfO_2$ . *Nano Lett.* **12**, 4318–4323 (2012).
33. Lakes, R. S., Lee, T., Bersie, A. & Wang, Y. C. Extreme damping in composite materials with negative-stiffness inclusions. *Nature* **410**, 565–567 (2001).
34. Jaglinski, T., Kochmann, D., Stone, D. & Lakes, R. S. Composite materials with viscoelastic stiffness greater than diamond. *Science* **315**, 620–622 (2007).
35. Ni, K. et al. Equivalent oxide thickness (EOT) scaling with hafnium zirconium oxide high- $\kappa$  dielectric near morphotropic phase boundary. In *2019 IEEE International Electron Devices Meeting* (IEDM) 7.4.1–7.4.4 (IEEE, 2019).
36. Budimir, M., Damjanovic, D. & Setter, N. Piezoelectric response and free-energy instability in the perovskite crystals  $BaTiO_3$ ,  $PbTiO_3$  and  $Pb(Zr, Ti)O_3$ . *Phys. Rev. B* **73**, 174106 (2006).
37. Noheda, B. et al. A monoclinic ferroelectric phase in the  $Pb(Zr_{1-x}Ti_x)O_3$  solid solution. *Appl. Phys. Lett.* **74**, 2059–2061 (1999).
38. Schroeder, U. et al. Recent progress for obtaining the ferroelectric phase in hafnium oxide based films: impact of oxygen and zirconium. *Jpn. J. Appl. Phys.* **58**, SL0801 (2019).
39. Schlom, D. G. & Haeni, J. H. A thermodynamic approach to selecting alternative gate dielectrics. *MRS Bull.* **27**, 198–204 (2002).
40. Bersuker, G. et al. The effect of interfacial layer properties on the performance of Hf-based gate stack devices. *J. Appl. Phys.* **100**, 094108 (2006).
41. Liao, Y.-H. et al. Electric field-induced permittivity enhancement in negative-capacitance FET. *IEEE Trans. Electron Devices* **68**, 1346–1351 (2021).

42. Ragnarsson, L.-Å. et al. Ultrathin EOT high- $\kappa$ /metal gate devices for future technologies: challenges, achievements and perspectives. *Microelectron. Eng.* **88**, 1317–1322 (2011).
43. Chatterjee, K., Rosner, A. J. & Salahuddin, S. Intrinsic speed limit of negative capacitance transistors. *IEEE Electron Device Lett.* **38**, 1328–1330 (2017).
44. Kwon, D. et al. Response speed of negative capacitance FinFETs. In *2018 IEEE Symposium on VLSI Technology* 49–50 (IEEE, 2018).
45. Pae, S. et al. Reliability characterization of 32 nm high-K and metal-gate logic transistor technology. In *2010 IEEE International Reliability Physics Symposium* 287–292 (IEEE, 2010).
46. Mukhopadhyay, S. et al. Trap generation in IL and HK layers during BTI/TDDB stress in scaled HKMG N and P MOSFETs and implications on  $t_{inv}$ -scaling. In *2014 IEEE International Reliability Physics Symposium* GD.3.1–GD.3.11 (IEEE, 2014).
47. Gao, W. et al. Room-temperature negative capacitance in a ferroelectric–dielectric superlattice heterostructure. *Nano Lett.* **14**, 5814–5819 (2014).
48. Zubko, P. et al. Negative capacitance in multidomain ferroelectric superlattices. *Nature* **534**, 524–528 (2016).
49. Wong, J. C. & Salahuddin, S. Negative capacitance transistors. *Proc. IEEE* **107**, 49–62 (2019).
50. Hsain, H. A., Lee, Y., Parsons, G. & Jones, J. L. Compositional dependence of crystallization temperatures and phase evolution in hafnia-zirconia  $(\text{Hf}_x\text{Zr}_{1-x})\text{O}_2$  thin films. *Appl. Phys. Lett.* **116**, 192901 (2020).
51. Lin, B.-T., Lu, Y.-W., Shieh, J. & Chen, M.-J. Induction of ferroelectricity in nanoscale  $\text{ZrO}_2$  thin films on Pt electrode without post-annealing. *J. Eur. Ceram. Soc.* **37**, 1135–1139 (2017).

52. Björck, M. & Andersson, G. GenX: an extensible X-ray reflectivity refinement program utilizing differential evolution. *J. Appl. Crystallogr.* **40**, 1174–1178 (2007).
53. Ilavsky, J. Nika: software for two-dimensional data reduction. *J. Appl. Crystallogr.* **45**, 324–328 (2012).
54. Park, M. H., Shimizu, T., Funakubo, H. & Schroeder, U. in *Ferroelectricity in Doped Hafnium Oxide: Materials, Properties and Devices* (eds Schroeder, U. et al.) 193–216 (Woodhead, 2019).
55. Mehmood, F., Mikolajick, T. & Schroeder, U. Lanthanum doping induced structural changes and their implications on ferroelectric properties of  $\text{Hf}_{1-x}\text{Zr}_x\text{O}_2$  thin film. *Appl. Phys. Lett.* **117**, 092902 (2020).
56. Mukundan, V. et al. Quantifying non-centrosymmetric orthorhombic phase fraction in 10 nm ferroelectric  $\text{Hf}_{0.5}\text{Zr}_{0.5}\text{O}_2$  films. *Appl. Phys. Lett.* **117**, 262905 (2020).
57. Park, M. H. et al. Ferroelectricity and antiferroelectricity of doped thin  $\text{HfO}_2$ -based films. *Adv. Mater.* **27**, 1811–1831 (2015).
58. Lyu, J., Fina, I., Solanas, R., Fontcuberta, J. & Sánchez, F. Growth window of ferroelectric epitaxial  $\text{Hf}_{0.5}\text{Zr}_{0.5}\text{O}_2$  thin films. *ACS Appl. Electron. Mater.* **1**, 220–228 (2019).
59. Young, A. T. et al. Variable linear polarization from an X-ray undulator. *J. Synchrotron Radiat.* **9**, 270–274 (2002).
60. Jain, A. et al. The Materials Project: a materials genome approach to accelerating materials innovation. *APL Mater.* **1**, 011002 (2013).
61. Mathew, K. et al. High-throughput computational X-ray absorption spectroscopy. *Sci. Data* **5**, 180151 (2018).
62. Cho, D.-Y., Jung, H.-S. & Hwang, C. S. Structural properties and electronic structure of  $\text{HfO}_2$ – $\text{ZrO}_2$  composite films. *Phys. Rev. B* **82**,

094104 (2010).

63. Park, M. H. & Hwang, C. S. Fluorite-structure antiferroelectrics. *Rep. Prog. Phys.* **82**, 124502 (2019).
64. Yang, K. & Hu, C. MOS capacitance measurements for high-leakage thin dielectrics. *IEEE Trans. Electron Devices* **46**, 1500–1501 (1999).
65. Changhwan C. et al. Fabrication of TaN-gated ultra-thin MOSFETS (EOT <1.0 nm) with HfO<sub>2</sub> using a novel oxygen scavenging process for sub 65 nm application. In *2005 Symposium on VLSI Technology* 226–227 (IEEE, 2005).
66. Takahashi, M. et al. Gate-first processed FUSI/HfO<sub>2</sub>/HfSiOx/Si MOSFETs with EOT=0.5 nm: interfacial layer formation by cycle-by-cycle deposition and annealing. In *2007 IEEE International Electron Devices Meeting (IEDM)* 523–526 (IEEE, 2007).
67. Mahapatra, S. (ed.) *Fundamentals of Bias Temperature Instability in MOS Transistors* (Springer, 2016).
68. Kim, Y. J. et al. Time-dependent negative capacitance effects in Al<sub>2</sub>O<sub>3</sub>/BaTiO<sub>3</sub> bilayers. *Nano Lett.* **16**, 4375–4381 (2016).
69. Hoffmann, M. et al. Demonstration of high-speed hysteresis-free negative capacitance in ferroelectric Hf<sub>0.5</sub>Zr<sub>0.5</sub>O<sub>2</sub>. In *2018 IEEE International Electron Devices Meeting (IEDM)* 31.6.1–31.6.4 (IEEE, 2018).
70. Kim, K. D. et al. Transient negative capacitance effect in atomic-layer-deposited Al<sub>2</sub>O<sub>3</sub>/Hf<sub>0.3</sub>Zr<sub>0.7</sub>O<sub>2</sub> bilayer thin film. *Adv. Funct. Mater.* **29**, 1808228 (2019).
71. Chen, L. Q. Phase-field method of phase transitions/domain structures in ferroelectric thin films: a review. *J. Am. Ceram. Soc.* **91**, 1835–1844 (2008).

72. Lomenzo, P. D. et al. A Gibbs energy view of double hysteresis in  $\text{ZrO}_2$  and Si-doped  $\text{HfO}_2$ . *Appl. Phys. Lett.* **117**, 142904 (2020).
73. Synopsys Sentaurus Device User Guide: Version O-2018.06 (Synopsys, 2018).
74. Park, J. Y. et al. A perspective on semiconductor devices based on fluorite-structured ferroelectrics from the materials–device integration perspective. *J. Appl. Phys.* **128**, 240904 (2020).
75. Hoffmann, M., Slesazeck, S., Schroeder, U. & Mikolajick, T. What's next for negative capacitance electronics? *Nat. Electron.* **3**, 504–506 (2020).
76. Hoffmann, M., Slesazeck, S. & Mikolajick, T. Progress and future prospects of negative capacitance electronics: a materials perspective. *APL Mater.* **9**, 020902 (2021).
77. Mikolajick, T. et al. Next generation ferroelectric materials for semiconductor process integration and their applications. *J. Appl. Phys.* **129**, 100901 (2021).
78. Li, Y.-L. et al. Electrical and reliability characteristics of FinFETs with high- $k$  gate stack and plasma treatments. *IEEE Trans. Electron Devices* **68**, 4–9 (2021).
79. Padmanabhan, R., Mohan, S., Morozumi, Y., Kaushal, S. & Bhat, N. Performance and reliability of  $\text{TiO}_2/\text{ZrO}_2/\text{TiO}_2$  (TZT) and AlO-doped TZT MIM capacitors. *IEEE Trans. Electron Devices* **63**, 3928–3935 (2016).
80. Shin, Y. et al. Crystallized HfLaO embedded tetragonal  $\text{ZrO}_2$  for dynamic random access memory capacitor dielectrics. *Appl. Phys. Lett.* **98**, 173505 (2011).
81. Mise, N. et al. Scalability of TiN/HfAlO/TiN MIM DRAM capacitor to 0.7-nm-EOT and beyond. In *2009 IEEE International Electron Devices Meeting (IEDM)* 11.3.1–11.3.4 (IEEE, 2009).

82. Kil, D.-S. et al. Development of new TiN/ZrO<sub>2</sub>/Al<sub>2</sub>O<sub>3</sub>/ZrO<sub>2</sub>/TiN capacitors extendable to 45nm generation DRAMs replacing HfO<sub>2</sub> based dielectrics. In *2006 Symposium on VLSI Technology* 38–39 (IEEE, 2006).
83. Kim, S. K. & Popovici, M. Future of dynamic random-access memory as main memory. *MRS Bull.* **43**, 334–339 (2018).
84. Park, M. H. et al. A comprehensive study on the mechanism of ferroelectric phase formation in hafnia-zirconia nanolaminates and superlattices. *Appl. Phys. Rev.* **6**, 041403 (2019).
85. Weeks, S. L., Pal, A., Narasimhan, V. K., Littau, K. A. & Chiang, T. Engineering of ferroelectric HfO<sub>2</sub>–ZrO<sub>2</sub> nanolaminates. *ACS Appl. Mater. Interfaces* **9**, 13440–13447 (2017).
86. Riedel, S., Polakowski, P. & Müller, J. A thermally robust and thickness independent ferroelectric phase in laminated hafnium zirconium oxide. *AIP Adv.* **6**, 095123 (2016).
87. Osada, M. & Sasaki, T. The rise of 2D dielectrics/ferroelectrics. *APL Mater.* **7**, 120902 (2019).
88. IRDS. Executive summary. In The International Roadmap for Devices and Systems: 2020 (IEEE, 2020); <http://irds.ieee.org>.
89. Park, H. W., Roh, J., Lee, Y. B. & Hwang, C. S. Modeling of negative capacitance in ferroelectric thin films. *Adv. Mater.* **31**, 1805266 (2019).
90. Park, M. H. et al. Morphotropic phase boundary of Hf<sub>1-x</sub>Zr<sub>x</sub>O<sub>2</sub> thin films for dynamic random access memories. *ACS Appl. Mater. Interfaces* **10**, 42666–42673 (2018).
91. Das, D. & Jeon, S. High- $\kappa$  Hf<sub>x</sub>Zr<sub>1-x</sub>O<sub>2</sub> ferroelectric insulator by utilizing high pressure anneal. *IEEE Trans. Electron Devices* **67**, 2489–2494 (2020).

92. Kim, S. et al. Method to achieve the morphotropic phase boundary in  $\text{Hf}_x\text{Zr}_{1-x}\text{O}_2$  by electric field cycling for DRAM cell capacitor applications. *IEEE Electron Device Lett.* **42**, 517–520 (2021).
93. Kashir, A. & Hwang, H. Ferroelectric and dielectric properties of  $\text{Hf}_{0.5}\text{Zr}_{0.5}\text{O}_2$  thin film near morphotropic phase boundary. *Phys. Status Solidi A* **218**, 2000819 (2021).
94. Appleby, D. J. R. et al. Experimental observation of negative capacitance in ferroelectrics at room temperature. *Nano Lett.* **14**, 3864–3868 (2014).

## Acknowledgements

This research was supported in part by the Berkeley Center for Negative Capacitance Transistors (BCNCT), the DARPA Technologies for Mixed-mode Ultra Scaled Integrated Circuits (T-MUSIC) programme, the University of California Multicampus Research Programs and Initiatives (UC MRPI) project and the US Department of Energy, Office of Science, Office of Basic Energy Sciences, Materials Sciences and Engineering Division under contract no. DE-AC02-05-CH11231 (Microelectronics Co-Design programme) for the development of materials for low-power microelectronics. This research used resources of the Advanced Photon Source, a US Department of Energy (DOE) Office of Science User Facility at Argonne National Laboratory and is based on research supported by the US DOE Office of Science, Basic Energy Sciences, under contract no. DE-AC02-06CH11357. V.A.S. and J.W.F. were supported by the US Department of Energy, Office of Science, Basic Energy Sciences, under award number DE-SC-0012375. Use of the Stanford Synchrotron Radiation Light source, 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. This research used resources of the Advanced Light Source, which is a DOE Office of Science User Facility under contract no. DE-AC02-05CH11231. Electron microscopy was performed at the Molecular Foundry, LBNL, supported by the Office of Science, Office of Basic Energy Sciences, US Department of

Energy (DE-AC02-05CH11231). Device fabrication was performed at the Marvell Nanofabrication Laboratory at UC Berkeley. This material is based upon work supported by the Secretary of Defense for Research and Engineering under Air Force contract no. FA8702-15-D-0001. Any opinions, findings, conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the Secretary of Defense for Research and Engineering.

## Author information

### Author notes

1. These authors contributed equally: Suraj S. Cheema, Nirmaan Shanker

### Affiliations

1. Department of Materials Science and Engineering, University of California, Berkeley, Berkeley, CA, USA

Suraj S. Cheema, Li-Chen Wang & Ramamoorthy Ramesh

2. Department of Electrical Engineering and Computer Sciences, University of California, Berkeley, Berkeley, CA, USA

Nirmaan Shanker, Cheng-Hsiang Hsu, Shang-Lin Hsu, Yu-Hung Liao, Wenshen Li, Jong-Ho Bae, Daewoong Kwon, Chenming Hu & Sayeef Salahuddin

3. Department of Electrical Engineering, University of Notre Dame, Notre Dame, IN, USA

Matthew San Jose, Jorge Gomez, Wriddhi Chakraborty, Patrick Fay & Suman Datta

4. Applied Science and Technology, University of California, Berkeley, CA, USA

Steve K. Volkman

5. Laser Thermal Laboratory, Department of Mechanical Engineering,  
University of California, Berkeley, CA, USA

Yoonsoo Rho & Costas P. Grigoropoulos

6. Lincoln Laboratory, Massachusetts Institute of Technology, Lexington,  
MA, USA

Gianni Pinelli, Ravi Rastogi, Dominick Pipitone, Corey Stull, Matthew  
Cook, Brian Tyrrell & Mohamed Mohamed

7. Department of Materials Science and Engineering, Pennsylvania State  
University, University Park, PA, USA

Vladimir A. Stoica

8. Advanced Photon Source, Argonne National Laboratory, Lemont, IL,  
USA

Zhan Zhang & John W. Freeland

9. Stanford Synchrotron Radiation Lightsource, SLAC National  
Accelerator Laboratory, Menlo Park, CA, USA

Christopher J. Tassone & Apurva Mehta

10. Applied Materials, Santa Clara, CA, USA

Ghazal Saheli & David Thompson

11. Research & Development Division, SK hynix, Icheon, Korea

Dong Ik Suh & Won-Tae Koo

12. Semiconductor R&D Center, Samsung Electronics, Gyeonggi-do,  
Korea

Kab-Jin Nam, Dong Jin Jung & Woo-Bin Song

13. Logic Technology Development, Intel Corporation, Hillsboro, OR, USA

Chung-Hsun Lin

14. Samsung Advanced Institute of Technology, Samsung Electronics, Suwon, Korea

Seunggeol Nam & Jinseong Heo

15. Department of Electrical Engineering, Indian Institute of Technology Bombay, Mumbai, India

Narendra Parihar & Souvik Mahapatra

16. Advanced Light Source, Lawrence Berkeley National Laboratory, Berkeley, CA, USA

Padraic Shafer

17. Department of Physics, University of California, Berkeley, Berkeley, CA, USA

Ramamoorthy Ramesh

18. Materials Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA, USA

Ramamoorthy Ramesh & Sayeef Salahuddin

19. National Center for Electron Microscopy, Molecular Foundry, Lawrence Berkeley National Laboratory, Berkeley, CA, USA

Jim Ciston

## Contributions

S.S.C. and S.S. designed the research. S.S.C. performed design, synthesis, and optimization of the superlattice oxide heterostructure and its ferroic characterization. N.S. and C.-H.H. performed capacitor fabrication. C.-H.H. helped optimize annealing treatments. N.S. and S.S.C. performed capacitor measurements and analysis. L.-C.W. fabricated the transistors and performed d.c. characterization. D.W.K. and J.B. developed the initial processes for transistor fabrication. Y.-H.L. performed simulations including EOT estimation, series resistance determination and mobility and transconductance analysis. M.S.J., J.G. and W.L. contributed to radio-frequency electrical measurements and analysis. N.S. performed MOSCAP stress measurements. W.C. and N.S. performed MOSFET reliability measurements under the guidance of S.D., S.M. and S.S. M.M., R.R., C.S., D.P., G.P., M.C. and B.T. contributed to capacitor fabrication and characterization at MIT LL. S.-L.H. performed TEM. S.S.C. and S.-L.H. performed TEM analysis. Y.R. performed second-harmonic generation. S.K.V. performed X-ray photoelectron spectroscopy. S.S.C. and C.-H.H. performed synchrotron soft X-ray spectroscopy at ALS. S.S.C., V.A.S. and J.W.F. performed synchrotron X-ray spectroscopy at APS. S.S.C., C.-H.H., V.A.S. and Z.Z. performed synchrotron in-plane diffraction at APS. S.S.C. and N.S. performed synchrotron X-ray reflectivity and 2D diffraction at SSRL. S.S.C. and S.S. co-wrote the original manuscript; S.S.C., N.S. and S.S. revised the manuscript. S.S. supervised the research. All authors contributed to discussions and manuscript preparations.

## Corresponding authors

Correspondence to [Suraj S. Cheema](#) or [Sayeef Salahuddin](#).

## Ethics declarations

## Competing interests

The authors declare no competing interests.

## Peer review

## Peer review information

*Nature* thanks Takashi Ando, Thomas Mikolajick and Burc Misirlioglu for their contribution to the peer review of this work.

## Additional information

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## Extended data figures and tables

### [Extended Data Fig. 1 Atomic-scale multilayer structure.](#)

**a**, Schematic of the HfO<sub>2</sub>–ZrO<sub>2</sub> multilayer structure on SiO<sub>2</sub>-buffered Si. **b**, Synchrotron X-ray reflectivity (XRR) of thicker HZH heterostructures (left) repeated with the same periodicity as the thinner trilayer structure; XRR fitting (right) demonstrates the presence of well separated HfO<sub>2</sub>–ZrO<sub>2</sub> layers, that is, not a solid solution, for three different multilayer repeats of fixed periodicity, all approximately following the expected 4 Å–12 Å–4 Å HZH structure. **c**, Layer-resolved electron energy loss spectroscopy (EELS) of the 2-nm HZH trilayer, demonstrating clear separation of HfO<sub>2</sub> and ZrO<sub>2</sub> layers. The exact layer thicknesses are extracted from XRR, which spans a wider sample footprint, rather than the local EELS measurement in which the apparent width increase can be due to beam spreading and local thickness variation. **d**, Angle-resolved X-ray photoelectric spectroscopy (XPS) of the 2-nm HZH trilayer (left) and the extracted atomic composition (right). The presence of increasing Zr content as the grazing angle increases is expected from the multilayer structure in which Zr content increases after the surface Hf-rich layer. Therefore XRR, EELS and XPS data all indicate the presence of a multilayer structure in which the HfO<sub>2</sub> layer is directly on the SiO<sub>2</sub> layer.

### [Extended Data Fig. 2 Ferroic phase insights from structural characterization.](#)

**a**, Left, in-plane synchrotron grazing-incidence diffraction (IP-GiD) of a bare 2-nm HZH trilayer indexed to the tetragonal P4<sub>2</sub>/nmc and orthorhombic Pca<sub>2</sub><sub>1</sub> phases. Right, magnification of the spectrum about the orthorhombic (111)<sub>o</sub> and tetragonal (101)<sub>t</sub> reflections, confirming the co-existing structural polymorphs in the 2-nm film. These two peaks were differentiated via self-consistent indexing of the entire spectrum, in which interplanar lattice spacings—determined from the {200}<sub>o</sub> family of reflections—closely match the *d* spacings for all other reflections—(111)<sub>o</sub>, (120)<sub>o</sub>, (211)<sub>o</sub>, (202)<sub>o</sub>—determined by Bragg’s law (Methods). **b**, Two-dimensional reciprocal space map of the bare 2-nm HZH trilayer, indexed by integrating the diffraction spectrum. The lack of fully polycrystalline rings illustrates that the 2-nm HZH trilayer is highly oriented, consistent with TEM imaging. **c**, Synchrotron spectroscopy (XAS) of the bare 2-nm HZH trilayer at the Hf M<sub>3</sub> edge (left) and Zr L<sub>3,2</sub> edge (centre); right, the presence of linear dichroism (orbital polarization) provides further evidence of symmetry-breaking in these oriented thin films. **d**, Second harmonic generation (SHG) mapped across the bare 2-nm HZH trilayer; the presence of SHG intensity confirms broken inversion symmetry in these ultrathin ferroic films. **e**, **f**, Additional cross-sectional TEM providing complementary evidence of the tetragonal P4<sub>2</sub>/nmc (**e**) and orthorhombic Pca<sub>2</sub><sub>1</sub> (**f**) phases, in which the extracted (101)<sub>t</sub> lattice spacing (~2.95 Å) and (111)<sub>o</sub> lattice spacing (~3.08 Å) extracted from IP-GiD are consistent with the average lattice spacings extracted from the periodicity of the TEM-imaged planes. The white scale bars in all the TEM images represent 1 nm.

### Extended Data Fig. 3 Ferroic phase insights: proximity to temperature-dependent phase transition.

**a**, Schematic of temperature-dependent AFE–FE phase evolution in fluorite-structure oxides. At lower temperatures, the higher symmetry tetragonal phase is expected to transition to the lower symmetry orthorhombic phase. **b**, Schematic crystal field splitting diagram for fluorite-structure polymorphs; the symmetry-induced *e*-splitting (rhombic distortion,  $\Delta_R$ ), besides the typical *t*<sub>2</sub>-*e* splitting (tetrahedral distortion,  $\Delta_T$ ) present in all fluorite-structure phases, provides a spectroscopic signature

for the polar O phase (Methods). **c**, Temperature-dependent XAS at the oxygen K edge for a 2-nm HZH bare film demonstrating clearer spectroscopic signatures of the FE O phase emerge slightly below room temperature. **d**, Simulated oxygen K-edge XAS spectra (Materials Project) for the respective O and T phases. XAS provides spectroscopic signatures to distinguish between the O and T phases (difficult to resolve from GI-XRD). **e**, Prototypical  $C-V$  behaviour for mixed AFE–FE (shoulder-like features in addition to the characteristic butterfly-like shape) and FE films (just butterfly-like) in MIM capacitor structures. **f**, Temperature-dependent  $C-V$  for thicker HZH multilayers of the same periodicity (in MIM capacitor structure) demonstrating an evolution from mixed-ferroic to FE-like hysteresis upon cooling slightly below room temperature. Thinner HZH multilayers films suffer from leakage limitations, preventing such hysteretic  $C-V$  measurements. The thicker HZH multilayers of the same periodicity—annealed at the same low-temperature condition to maintain the multilayer structure—demonstrate a similar mixed ferroic to FE phase transition slightly below room temperature as the thinner 2-nm multilayer (c).

#### **Extended Data Fig. 4 Solid solutions versus superlattice structure: role of ALD period and Zr content.**

**a**, Schematic of HZH multilayer and Zr-rich  $\text{Hf}:\text{ZrO}_2$  solid-solution films. With shorter ALD periods, the mixed FE–AFE multilayer structure transitions towards a  $\text{Hf}:\text{ZrO}_2$  solid solution with AFE-like behaviour. In the solid-solution state, the loss of the mixed ferroic order yields diminished capacitance, owing to the lack of mixed-ferroic-induced capacitance enhancement (Fig. 1a). **b**, MOS accumulation  $C-V$  of the HZH trilayer (60% Zr) compared to solid-solution films of the same thickness (2 nm) and composition (60% Zr), as well as solid-solution films of the same thickness and higher Zr composition (67%–100% Zr). **c**, MIM  $C-V$  hysteresis loops of the HZH superlattice (60% Zr) compared to solid-solution films of the same thickness (6 nm) and composition (60% Zr), as well as solid-solution films of the same thickness and higher Zr composition (67%–100% Zr).  $\text{Hf}:\text{ZrO}_2$  solid-solution films with higher Zr content (60%–75%) are around the range attributed to the ‘MPB’ in thicker  $\text{Hf}:\text{ZrO}_2$  alloys<sup>35,55,90,91,92,93</sup>. These results indicate that the capacitance enhancement in multilayer films

is not simply driven by Zr content<sup>32,38,57,63</sup>, but instead the atomic-scale stacking, as the solid-solution films with subatomic superlattice period do not demonstrate the same mixed-ferroic behaviour and enhanced capacitance as the superlattices.

### **Extended Data Fig. 5 Solid solutions versus superlattice structure: role of annealing temperature.**

**a**, Schematic of  $\text{HfO}_2$ – $\text{ZrO}_2$  multilayer and  $\text{Hf}:\text{ZrO}_2$  solid-solution films. Under a high-temperature anneal, the multilayer structure transitions towards a  $\text{Hf}:\text{ZrO}_2$  solid-solution-like structure demonstrating more FE-like behaviour. The solid-solution state yields diminished capacitance owing to the lack of both the higher-permittivity AFE phase and the mixed-ferroic-induced capacitance enhancement (Fig. 1a). **b**, Comparison of MOS capacitor accumulation  $C$ – $V$  characteristics in HZH multilayers, where the superstructure was repeated (left) one, (centre) two, or (right) three times, under both low- and high-temperature anneals. **c**, Comparison of mixed-ferroic behaviour in low-temperature treated MIM HZH multilayers versus FE behaviour in the same multilayers annealed at high temperatures, where the superstructure was repeated (left) three, (centre) four, or (right) five times. In all instances, the high-temperature anneal ( $>500$  °C) results in diminished accumulation capacitance compared to the low-temperature anneals, as the multilayered mixed-ferroic films presumably transition to more FE-like solid-solution alloys.

### **Extended Data Fig. 6 $\text{SiO}_2$ interlayer thickness.**

**a**, Wide field-of-view (FOV) cross-sectional TEM images of the HZH multilayer structure and its corresponding intensity line scan (bottom right) averaged across the entire top cross-sectional image FOV (~150 nm, teal-coloured box). Note the vertical teal-coloured lines in the intensity line scan correspond to the inner teal-coloured box in the wide-FOV image, which delineate the  $\text{SiO}_2$  interlayer boundaries. The bottom cross-sectional TEM image highlights the thin  $\text{SiO}_2$  interlayer (white region) without obfuscation by the teal-coloured box. A physical  $\text{SiO}_2$  thickness of 8.6 Å is extracted

from analysis of the averaged intensity line scan of the wide FOV TEM (Methods). **b**, **d**,  $C-V$  measurements of  $\text{HfO}_2$  (**b**) and  $\text{Al}_2\text{O}_3$  (**d**) thickness series in MOS capacitor structures (left), extracted inverse capacitance versus thickness at different charge values (centre), and extracted  $Q-V$  relation Si charge layer and  $\text{SiO}_2$  interlayer (SiL; right), which fits to TCAD simulations for 8.0 Å  $\text{SiO}_2$ . The SiL  $Q-V$  relation was found by integrating the extracted capacitance equivalent thickness of SiL versus charge (right, inset). This electrical interlayer thickness (8.0 Å) is slightly less than the physical thickness determined by TEM (8.6 Å). As a sanity check, the extracted permittivity from this methodology for  $\text{HfO}_2$  and  $\text{Al}_2\text{O}_3$  corresponds to 19 and 9, respectively, as is expected (Methods section ‘Permittivity extraction’). **c**, **e**,  $Q-V$  curves for  $\text{HfO}_2$  (**c**) and  $\text{Al}_2\text{O}_3$  (**e**) thickness series obtained from integrating MOS  $C-V$  measurements (left), extracted voltage versus thickness at various charge values (centre), and extracted  $Q-V$  relation of SiL (right). The SiL  $Q-V$  relation is consistent with the  $Q-V$  relation extracted from the  $C-V$  data (inset). **f**, Consistency in the SiL  $Q-V$  relation extracted from the  $C-V$  data from both the  $\text{HfO}_2$  and  $\text{Al}_2\text{O}_3$  thickness series, which both fit to 8.0 Å  $\text{SiO}_2$  interlayer thickness.

## Extended Data Fig. 7 Capacitance and charge enhancement.

**a**, MOS schematic of the 20 Å HZH mixed-ferroic trilayer sample on lightly doped Si ( $10^{15} \text{ cm}^{-3}$ ) considered for the following  $C-V$  and pulsed  $I-V$  measurements. **b**, Accumulation  $C-V$  curves for 2-nm HZH grown on sub-nm  $\text{SiO}_2$  fit to equivalent oxide thickness (EOT) simulations (Methods). Inset, externally verified MOS accumulation  $C-V$  of the same trilayer stack (Methods), demonstrating 6.5 Å EOT. The 2-nm trilayer on top of  $\text{SiO}_2$  demonstrates lower EOT than the thickness of  $\text{SiO}_2$  interlayer alone, carefully extracted via physical (8.5 Å) and electrical (8.0 Å) methodologies (Extended Data Fig. 6), providing evidence of capacitance enhancement. **c**, The applied voltage pulse (top), measured current response (centre), and integrated charge (bottom) as a function of time for 2-nm HZH in MOS capacitors. **d**, The maximum charge  $Q_{\max}$ , the residual charge  $Q_{\text{res}}$ , and their difference,  $Q_{\text{rev}}$ , derived from the charge versus time curve for each of the voltage pulses (Methods). **e**, The reversible charge of the MOS layer

(top) compared against the extracted charge of the Si charge layer plus  $\text{SiO}_2$  interlayer (SiL) derived electrically (Extended Data Fig. 6f). The charge boost (bottom) present in the total MOS structure (SiL plus HZH capacitors) compared to just the SiL is a signature of negative capacitance<sup>26,69</sup>. f, The polarization–electric field ( $P-E_F$ ) relationship for just the 2-nm HZH layer, extracted from the charge–voltage relationship in e. Note that the presence of a negative slope regime in the extracted  $P-E_F$  relation corresponds to negative capacitance stabilization<sup>26,69</sup>. g, Scatter plot of reported FE–dielectric systems demonstrating capacitance or charge enhancement at the capacitor-level, via  $C-V$  or pulsed  $I-V$  measurements, respectively. The plot considers fluorite-structure bilayers<sup>26,69</sup> (red), perovskite-structure bilayers<sup>29,94</sup> (blue, BL), and perovskite-structure superlattices<sup>30,31,47,48</sup> (blue, SL). This work marks, to our knowledge, the thinnest demonstration of negative capacitance.

## Extended Data Fig. 8 Mobility and reliability.

a, Intrinsic  $C_{gg}$  versus  $V_{gs} - V_{fb}$  for  $\sim 20 \text{ \AA}$  HZH and  $\text{HfO}_2$  gate stacks, which fit to  $7.5 \text{ \AA}$  and  $9.5 \text{ \AA}$ , respectively, extracted from SOI transistors. b,  $I_d$  versus  $V_{gs} - V_T$  at  $V_d = 50 \text{ mV}$  for transistors implementing HZH and  $\text{HfO}_2$  gate stacks. c, EOT and mobility (at  $5 \times 10^{12} \text{ cm}^{-2}$ ) for HZH and  $\text{HfO}_2$ , demonstrating no mobility degradation. d, Normalized mobility versus EOT for transistors integrating the 2-nm HZH mixed-ferroic gate stack (blue) versus a 2-nm  $\text{HfO}_2$  standard high- $\kappa$  dielectric gate stack (black) of higher EOT, demonstrating no mobility degradation. These results are also benchmarked against reported HKMG literature<sup>3</sup> implementing interlayer-scavenged 2-nm  $\text{HfO}_2$  (red). Inset,  $\text{SiO}_2$  interlayer thickness versus EOT for  $6.5 \text{ \AA}$  EOT HZH stack against notable HKMG literature employ interlayer scavenging<sup>3</sup>. This scatter plot highlights the underlying reason for the improved leakage-EOT and mobility-EOT behaviour in the ultrathin HZH gate stacks: achieving low EOT without reducing the  $\text{SiO}_2$  interlayer thickness. e, Measure–stress–measure scheme used in PBTI reliability measurements. f, g,  $\Delta V_T$  versus stress time for long-channel bulk transistors integrating HZH (f) and  $\text{HfO}_2$  (g) at varying stress

conditions (up to 9 MV cm<sup>-1</sup>), measured at 85 °C. The extracted time exponent,  $n$ , is similar to reported high- $\kappa$  HfO<sub>2</sub> stacks<sup>67</sup>, which is expected considering the similar interfacial (IL) oxide and IL-high- $\kappa$  interface<sup>46</sup>; both stacks with different EOT have HfO<sub>2</sub> sitting on the same SiO<sub>2</sub> IL (Extended Data Fig. 1). **h**, d.c. lifetime (stress time needed to induce a 50 mV  $\Delta V_T$  shift) versus electric field for HZH and HfO<sub>2</sub>, demonstrating similar rates of degradation. **i, j**, Evolution of MOSCAP accumulation  $C$ – $V$  curves as a function of stress time at a stress of  $V_{fb} - 1$  V for HZH (**i**) and HfO<sub>2</sub> (**j**).

### Extended Data Fig. 9 Radio frequency device characterization.

**a**, De-embedding procedure for extracting corrected admittance parameters ( $Y_{corr}$ ) by decoupling parasitic shunt capacitance and series resistance and inductance by measuring scattering parameters for the device under test (DUT) as well as open and short structures. More details can be found in Methods. **b**, Small-signal model for transistor used to extract transconductance ( $g_m$ ) and total gate capacitance ( $C_{gg} = C_{gs} + C_{gd}$ ). **c**, De-embedded  $(2\pi f)^{-1}\text{Re}(Y_{21})$  points extrapolated to the zero frequency limit (dotted lines) to extract the radio frequency  $g_m$ . All data shown were extracted from bulk transistors ( $L_G = 1 \mu\text{m}$ ) integrating the 2-nm HZH ferroic gate stack.

### Extended Data Fig. 10 Transconductance extraction.

**a**, Threshold voltage extraction by linear extrapolation for various channel lengths. All channel lengths give nearly constant  $V_T$  (~0.42 V), satisfying the assumption for the line resistance method. **b**, Source/drain series resistance extracted using the  $1/V_{ov}$  method (Methods). By performing a linear interpolation of the total resistance for  $V_{ov} = 0.5$ –0.6 V, the extracted series resistance is ~500 Ω μm. **c**, Source/drain series resistance extracted using the line resistance method (Methods). The trend is considered down to  $L_G = 90$  nm, which intersects at ~500–600 Ω μm—consistent with the  $1/V_{ov}$  method—with an  $L_G$  offset of ~50 nm. **d, e**, Measured (left) and extracted (right) transconductance (**d**) and output conductance (**e**) versus  $V_g$

for  $V_{ds} = 0.9\text{--}1.1$  V, assuming  $R_s = R_d = 250 \Omega \mu\text{m}$  for  $L_G = 90$  nm. The deembedding of intrinsic  $g_{m,i}$  and  $g_{ds,i}$  from extrinsic  $g_m$  and  $g_{ds}$  is described in Methods. All data shown were measured on SOI short-channel transistors integrating the 2-nm HZH ferroic gate stack.

## Rights and permissions

### Reprints and Permissions

## About this article

### Cite this article

Cheema, S.S., Shanker, N., Wang, LC. *et al.* Ultrathin ferroic  $\text{HfO}_2\text{--ZrO}_2$  superlattice gate stack for advanced transistors. *Nature* **604**, 65–71 (2022). <https://doi.org/10.1038/s41586-022-04425-6>

- Received: 11 April 2021
- Accepted: 14 January 2022
- Published: 06 April 2022
- Issue Date: 07 April 2022
- DOI: <https://doi.org/10.1038/s41586-022-04425-6>

## Share this article

Anyone you share the following link with will be able to read this content:

[Get shareable link](#)

Sorry, a shareable link is not currently available for this article.

[Copy to clipboard](#)

Provided by the Springer Nature SharedIt content-sharing initiative

---

This article was downloaded by **calibre** from <https://www.nature.com/articles/s41586-022-04425-6>

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

[Download PDF](#)

- Article
- Open Access
- [Published: 06 April 2022](#)

# Reconstructed covalent organic frameworks

- [Weiwei Zhang](#)<sup>1</sup>,
- [Linjiang Chen](#)<sup>1,2</sup>,
- [Sheng Dai](#) ORCID: [orcid.org/0000-0001-5787-0179](https://orcid.org/0000-0001-5787-0179)<sup>1</sup>,
- [Chengxi Zhao](#)<sup>1,2</sup>,
- [Cheng Ma](#)<sup>3</sup>,
- [Lei Wei](#)<sup>4</sup>,
- [Minghui Zhu](#) ORCID: [orcid.org/0000-0003-1593-9320](https://orcid.org/0000-0003-1593-9320)<sup>3</sup>,
- [Samantha Y. Chong](#) ORCID: [orcid.org/0000-0002-3095-875X](https://orcid.org/0000-0002-3095-875X)<sup>2</sup>,
- [Haofan Yang](#)<sup>2</sup>,
- [Lunjie Liu](#)<sup>2</sup>,
- [Yang Bai](#) ORCID: [orcid.org/0000-0002-1643-3770](https://orcid.org/0000-0002-1643-3770)<sup>2</sup>,
- [Miaojie Yu](#)<sup>1</sup>,
- [Yongjie Xu](#)<sup>2</sup>,
- [Xiao-Wei Zhu](#) ORCID: [orcid.org/0000-0002-9778-9233](https://orcid.org/0000-0002-9778-9233)<sup>2</sup>,
- [Qiang Zhu](#)<sup>2</sup>,
- [Shuhao An](#)<sup>1</sup>,
- [Reiner Sebastian Sprick](#) ORCID: [orcid.org/0000-0002-5389-2706](https://orcid.org/0000-0002-5389-2706)<sup>2</sup>,
- [Marc A. Little](#) ORCID: [orcid.org/0000-0002-1994-0591](https://orcid.org/0000-0002-1994-0591)<sup>2</sup>,
- [Xiaofeng Wu](#) ORCID: [orcid.org/0000-0001-5549-8836](https://orcid.org/0000-0001-5549-8836)<sup>1,2</sup>,
- [Shan Jiang](#)<sup>4</sup>,
- [Yongzhen Wu](#)<sup>1</sup>,
- [Yue-Biao Zhang](#) ORCID: [orcid.org/0000-0002-8270-1067](https://orcid.org/0000-0002-8270-1067)<sup>4</sup>,
- [He Tian](#) ORCID: [orcid.org/0000-0003-3547-7485](https://orcid.org/0000-0003-3547-7485)<sup>1</sup>,
- [Wei-Hong Zhu](#) ORCID: [orcid.org/0000-0001-9103-166X](https://orcid.org/0000-0001-9103-166X)<sup>1</sup> &
- [Andrew I. Cooper](#) ORCID: [orcid.org/0000-0003-0201-1021](https://orcid.org/0000-0003-0201-1021)<sup>1,2</sup>

*Nature* volume 604, pages 72–79 (2022)

- 8384 Accesses
- 60 Altmetric
- [Metrics details](#)

## Subjects

- [Organic molecules in materials science](#)
- [Photocatalysis](#)
- [Polymer synthesis](#)
- [Porous materials](#)

## Abstract

Covalent organic frameworks (COFs) are distinguished from other organic polymers by their crystallinity<sup>1,2,3</sup>, but it remains challenging to obtain robust, highly crystalline COFs because the framework-forming reactions are poorly reversible<sup>4,5</sup>. More reversible chemistry can improve crystallinity<sup>6,7,8,9</sup>, but this typically yields COFs with poor physicochemical stability and limited application scope<sup>5</sup>. Here we report a general and scalable protocol to prepare robust, highly crystalline imine COFs, based on an unexpected framework reconstruction. In contrast to standard approaches in which monomers are initially randomly aligned, our method involves the pre-organization of monomers using a reversible and removable covalent tether, followed by confined polymerization. This reconstruction route produces reconstructed COFs with greatly enhanced crystallinity and much higher porosity by means of a simple vacuum-free synthetic procedure. The increased crystallinity in the reconstructed COFs improves charge carrier transport, leading to sacrificial photocatalytic hydrogen evolution rates of up to 27.98 mmol h<sup>-1</sup> g<sup>-1</sup>. This nanoconfinement-assisted reconstruction strategy is a step towards programming function in organic materials through atomistic structural control.

## Main

Covalent organic frameworks (COFs) are of growing interest for gas storage, separation, electronics and catalysis applications because of their predictable structures and ordered nanopores<sup>10,11,12,13,14,15,16,17,18,19</sup>. Two-dimensional COFs with π-stacking between the layers allow for charge carrier transport in aligned molecular columns, and these materials show promise for photoenergy conversion and optoelectronics<sup>20,21,22,23,24,25,26,27,28,29,30,31</sup>. However, material quality and sometimes demanding synthetic procedures can limit practical applications. In particular, the

moderate level of crystallinity in two-dimensional COFs can compromise their performance in optoelectronic applications, and synthetic requirements such as vacuum sealing or strictly anaerobic conditions are practical hurdles to scale-up.

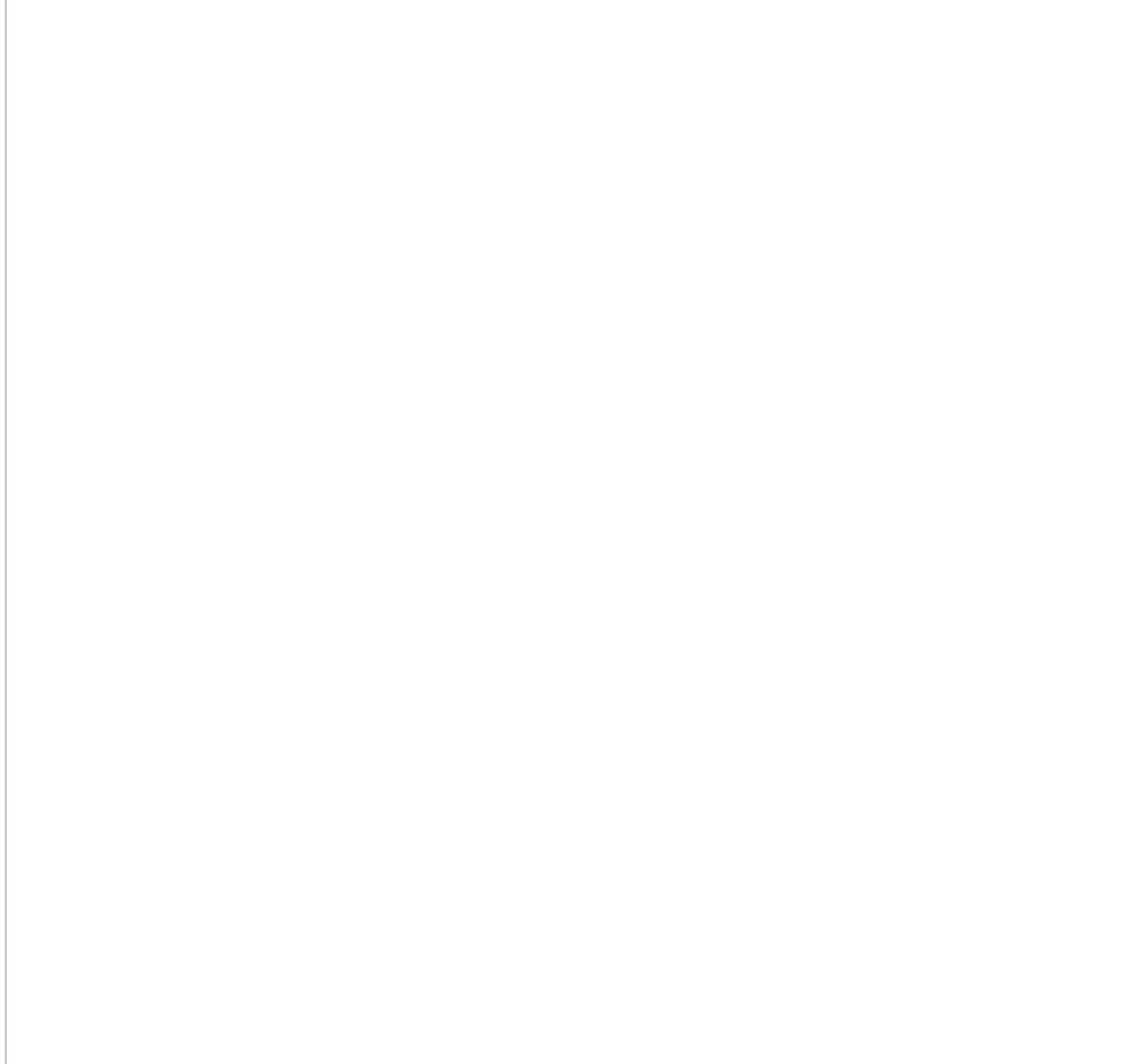
COFs are typically prepared by simultaneous polymerization and crystallization of monomers following the principle of dynamic covalent chemistry<sup>32</sup>. Reversible bond formation and structural self-healing have a central role in achieving long-range crystalline order. Strategies have been reported to produce COFs using more reversible chemistry, even to the point of obtaining single crystals<sup>6,7,8,9</sup>, but not all of these strategies lead to porosity. High degrees of crystallinity are much more difficult to obtain when the framework bonding is more robust and less reversible<sup>4,5,33</sup>. As such, there is a trade-off between stability—which is desirable for practical applications—and high levels of long-range crystalline order. An attractive strategy is to pre-organize monomers before polymerization. This separates the crystallization process from the (irreversible) bond forming step<sup>34,35,36,37,38,39</sup>. In this case, monomers are pre-arranged in the solid state to form an ordered self-assembled structure before the polymerization reaction. However, the pre-organization is often based on weak intermolecular interactions, and strain induced by the change in geometry that occurs during polymerization can cause fragmentation of crystallites or structural disorder. Thus, such reactions tend to be limited to mild photo-polymerizations in which the change in geometry is not too large<sup>34,35,36</sup>, although more profound structural transformations are possible when flat and rigid building blocks are used<sup>37,38</sup>. The success of these solid-state transformation strategies relies on appropriate molecular ordering, which is hard to design *a priori*: for example, pre-organized non-covalent molecular crystals can form different polymorphs in different crystallization solvents, which can thwart reticular framework strategies.

Confinement effects are ubiquitous in the chemistry of life; they prevent the denaturation of proteins and allow the synthesis of complex biomolecules under mild conditions. Likewise in synthetic nanochemistry, confining molecules can profoundly affect reaction pathways by stabilizing reactive species, accelerating reactions or enhancing selectivity<sup>40,41</sup>. Here we present a reconstruction strategy for COF synthesis that uses a reversible and removable covalent tether to pre-organize monomers before an irreversible polymerization. This route yields highly crystalline and functional COF materials through a facile process (Fig. 1a). By stepwise control over temperature and solvent, we achieved a chemical reconstruction in pre-organized urea-linked COFs. Instead of becoming amorphous, solvothermal treatment initiates a multi-step urea hydrolysis reaction followed by imine condensation. Notably, this generates a highly crystalline reconstructed COF (RC-COF) through a framework transformation, even though the mass loss during reconstruction can be as high as 36%. The position of the monomers that are produced by hydrolysis is directed by nanoconfinement in the framework before *in situ* polymerization. This results in greatly improved crystallinity

and functional properties for the RC-COFs compared to directly polymerized imine frameworks, in which the monomers are aligned randomly before polymerization.

**Fig. 1: Chemical reconstruction.**

 figure 1



**a**, The synthetic procedure for reconstructed COFs includes two steps: pre-organization of the monomers using reversible urea linkages to form a highly crystalline framework, followed by a solvothermal treatment step that removes the urea tethers, releasing monomers that then undergo *in situ* polymerization to form the reconstructed  $\beta$ -ketoenamine COF. The urea linkage acts as a disposable tether in this one-pot multi-step reaction, organizing the monomers before being removed by conversion into ammonia and carbon dioxide gas. **b**, Transformation of the model compound. A small urea-linked model compound can also be converted into the corresponding  $\beta$ -ketoenamine product, but with low isolated yield (around 11% yield)

in the solid state in the presence of H<sub>2</sub>O; decomposition occurs when the model compound is dissolved in solution, NMP/H<sub>2</sub>O (9/1 v/v), and the β-ketoenamine product is not detected).

## Structural transformation

Urea chemistry is inexpensive and used in the large-scale manufacture of resins and adhesives. Urea-linked COFs have been synthesized previously<sup>42</sup>. Urea is quite stable, with a decomposition half-life of 3.6 years in aqueous solution (38 °C); in industry, hydrolysis of urea feedstocks into ammonia and carbon dioxide is used for ammonia supply. This hydrolysis reaction is favoured by increased temperatures. We therefore speculated that raising the temperature could promote a structural change in urea-linked COFs. Urea-COF-1 (also known as COF-117; ref. <sup>42</sup>) was first synthesized via a Schiff-base condensation reaction of 1,1'-(1,4-phenylene)diurea with 1,3,5-triformylphloroglucinol in a mixture of *N*-methyl-2-pyrrolidinone (NMP), 1,2-dichlorobenzene (*o*-DCB) and aqueous acetic acid (6 mol l<sup>-1</sup>) at 90 °C for 72 h. Instead of isolating the powdered urea COF, we directly raised the temperature to 110, 120, 130, 150, 160 and 170 °C, respectively, for a further 72 h. A colour change from yellow to dark red was observed after raising the temperature (Fig. [2a](#)), suggesting more extended electronic conjugation. The crystallinity of the solvated samples was assessed by powder X-ray diffraction (PXRD; Fig. [2a](#)). When the reaction temperature was increased from 90 °C to 160 °C, the first intense diffraction peak was found to shift gradually from  $2\theta = 3.5^\circ$  to  $4.6^\circ$ . An increase in the reaction temperature to 170 °C did not shift this peak any further (Supplementary Fig. [3](#)).

**Fig. 2: Thermal and water-triggered reconstruction.**

---

 **figure 2**

**a, b**, Evolution of the PXRD patterns (**a**) and FTIR spectra (**b**) for Urea-COF-1 as-synthesized at 90 °C and treated at the in situ increased reaction temperatures of 110, 120, 130, 150 and 160 °C for a further 72 h, respectively, in an 8/2/1 mixture of NMP, *o*-DCB and 6 mol l<sup>-1</sup> acetic acid. The gradual shift in PXRD peak positions suggests a continuous structural transformation. Insets are photographs of isolated powders. AU, arbitrary units. **c, d**, Evolution of the PXRD patterns (**c**) and FTIR spectra (**d**) for isolated Urea-COF-1 treated by solvents with increased water content (*o*-DCB, NMP,

glacial acetic acid, NMP/H<sub>2</sub>O (9/1 v/v) and H<sub>2</sub>O, respectively) at 160 °C for 72 h. Insets are photographs of Urea-COF-1 (yellow) and RC-COF-1 (dark red) powders. A comparison of the FTIR spectrum with that of DP-COF-1, which was synthesized by direct imine polycondensation, is also shown in **d**. **e**, <sup>13</sup>C CP-MAS solid-state NMR spectra of Urea-COF-1, RC-COF-1 and DP-COF-1. Spinning sidebands are denoted with asterisks. Carbon atoms responsible for the NMR resonances are labelled A–F (for DP-COF-1 and RC-COF-1) and a–f (for Urea-COF-1).

### Source data

Various solvents and solvent mixtures were investigated for this solvothermal treatment, such as *o*-DCB, NMP, glacial acetic acid and water. We first isolated the Urea-COF-1 powders formed at 90 °C after 72 h, and then used these solvents or solvent mixtures to treat the urea COF in a sealed Pyrex tube, separately, at a fixed temperature of 160 °C for a further 72 h (Fig. [2c](#)). Urea-COF-1 decomposed in neat *o*-DCB and NMP, and little solid material was isolated after thermal treatment. In glacial acetic acid, most of the solid was retained and the first diffraction peak shifted to 4.4°, suggesting an incomplete phase transformation. When Urea-COF-1 was treated with pure water at 160 °C (Fig. [2c](#), Extended Data Fig. [1a](#)), an intense diffraction peak appeared at  $2\theta = 4.6^\circ$  and the colour of the powder changed from yellow to dark red (RC-COF-1). A small quantity of water in NMP (NMP/H<sub>2</sub>O; 9/1 v/v) also promoted this reconstruction. These experiments demonstrate that the transformation was induced by water, along with the increased temperature. A high concentration of ammonium ion was detected from the aqueous solution (Supplementary Fig. [4](#)). Elemental analysis showed a marked reduction in the nitrogen content for RC-COF-1 relative to Urea-COF-1 (11.89 versus 16.75 wt%), and the experimental weight loss during transformation (36 wt%) was close to the proportion of urea in Urea-COF-1 (theoretical mass loss = 29 wt%), ignoring any end groups. We therefore hypothesized that the urea-linked COF had transformed into a  $\beta$ -ketoenamine COF by solvothermal treatment in water, with ammonia and carbon dioxide being released as by-products (Fig. [1](#)). RC-COF-1 retained high crystallinity after thermal desolvation under dynamic vacuum, in contrast to Urea-COF-1, which has flexible linkages and loses crystallinity after solvent removal<sup>[42](#)</sup> (Extended Data Fig. [1b](#)).

To further confirm the structure of RC-COF-1, we synthesized the same  $\beta$ -ketoenamine COF by direct polymerization (DP-COF-1; also known as TpPa-1<sup>[4](#)</sup>) of 1,3,5-triformylphloroglucinol with *p*-phenylenediamine according to reported procedures. <sup>13</sup>C cross-polarization magic angle spinning (CP-MAS) solid-state nuclear magnetic resonance (NMR) spectroscopy showed the same resonances for RC-COF-1 and DP-COF-1, although the peaks were narrower and better resolved in RC-COF-1, suggesting increased structural order<sup>[43](#)</sup> (Fig. [2e](#)). Fourier transform infrared (FTIR) spectroscopy for activated Urea-COF-1 showed strong bands at around 1,713 and

$3,285\text{ cm}^{-1}$ , corresponding to urea C=O and N–H groups (Fig. 2d). These bands disappeared after increasing the reaction temperature to 110–160 °C (Fig. 2b); this change was even more noticeable when we increased the water content in the solvent (Fig. 2d), suggesting hydrolysis of the urea groups. Elemental analysis (Supplementary Table 1) and X-ray photoelectron spectroscopy (Supplementary Fig. 6) also supported the solvothermal transformation of Urea-COF-1 to the  $\beta$ -ketoenamine COF, RC-COF-1.

## Improved crystallinity and surface area

The level of crystallinity in RC-COF-1 was markedly enhanced compared to its directly polymerized analogue, DP-COF-1 (Fig. 3c); RC-COF-1 showed prominent diffraction peaks at 4.6, 8.1, 9.3, 12.3, 14.0, 16.1, 16.6, 18.7, 20.3 and 27.1°, which were indexed as 100, 110, 200, 120, 300, 220, 130, 400, 410 and 001 reflections, respectively. By contrast, only four broad peaks could be discerned from the diffraction pattern of DP-COF-1, as measured using the same diffraction set-up and measurement conditions. Eclipsed AA-stacking models yielded PXRD patterns that were consistent with the experimental profiles of Urea-COF-1 and RC-COF-1 (Supplementary Figs. 9, 17). Pawley refinement in the  $P6/m$  space group with unit cell parameters of  $a = 29.39$ ,  $b = 29.39$ ,  $c = 3.56\text{ \AA}$  (Fig. 3a) and  $a = 22.04$ ,  $b = 22.04$ ,  $c = 3.49\text{ \AA}$  (Fig. 3b) reproduced the experimental curve with good agreement factors (weight-profile  $R$ -factor  $R_{wp} = 5.52\%$  and unweighted  $R$ -factor  $R_p = 4.34\%$  for Urea-COF-1, and  $R_{wp} = 4.64\%$  and  $R_p = 3.36\%$  for RC-COF-1), which suggested a pronounced contraction of the unit cell after reconstruction.

**Fig. 3: Reconstructed COFs with enhanced crystallinity and porosity.**

---

 figure 3

**a, b**, Simulated and experimental PXRD patterns for Urea-COF-1 (solvated) (**a**) and RC-COF-1 (activated) (**b**). The structural models were built using Materials Studio and refined using experimental PXRD data. **c**, Comparison of PXRD patterns for RC-COF-1 synthesized by the reconstruction protocol and DP-COF-1 synthesized by

direct polymerization. **d**, Nitrogen adsorption isotherm (filled symbols) and desorption isotherm (open symbols) for RC-COF-1, DP-COF-1 and Urea-COF-1 recorded at 77.3 K; RC-COF-1 shows a type I isotherm. **e**, SEM image of RC-COF-1. Scale bar, 1  $\mu$ m. **f**, HRTEM image of RC-COF-1. Insets show the FFT pattern taken from the regions highlighted by the dashed-line squares and the corresponding filtered inverse FFT image. Scale bars, 50 nm (main image); 10 nm (inset). **g–i**, Simulated and experimental PXRD patterns for RC-COF-2 (**g**), RC-COF-3 (**h**) and RC-COF-4 (**i**), and comparison with PXRD patterns of DP-COF-2, DP-COF-3 and DP-COF-4 synthesized by direct polymerization. **j**, Nitrogen adsorption isotherm (filled symbols) and desorption isotherm (open symbols) for RC-COF-2, RC-COF-3, RC-COF-4 and directly polymerized analogues. **k**, **l**, HRTEM images of RC-COF-2 (**k**) and RC-COF-3 (**l**). Insets show FFT patterns and the corresponding filtered inverse FFT images. Scale bars, 50 nm (main images); 10 nm (insets).

### Source data

The porosity of these COFs was evaluated by nitrogen adsorption measurements at 77.3 K (Fig. 3d). Urea-COF-1 showed a low Brunauer–Emmett–Teller (BET) surface area of  $38 \text{ m}^2 \text{ g}^{-1}$  because of pore deformation upon activation<sup>42</sup>. This increased to  $1,712 \text{ m}^2 \text{ g}^{-1}$  for RC-COF-1, which showed a type I gas adsorption isotherm with rapid gas uptake at low relative pressures ( $P/P_0 < 0.01$ ), indicating a highly microporous solid (Fig. 3d, Extended Data Fig. 2). The narrow pore size distribution of around 1.6 nm obtained from the adsorption isotherm using nonlocal density functional theory fitting was in precise agreement with the proposed structural model (Extended Data Fig. 3). By contrast, DP-COF-1 adsorbed much less gas (Fig. 3d) and showed a much lower BET surface area of  $580 \text{ m}^2 \text{ g}^{-1}$ , close to previous reports for this material<sup>4</sup> ( $535 \text{ m}^2 \text{ g}^{-1}$ ). The broader, less regular pore size distribution for DP-COF-1 (Extended Data Fig. 3) can be ascribed to its semi-crystalline nature. The high crystallinity and regular porosity of RC-COF-1 also translated into high  $\text{CO}_2$  uptake, as shown by gas adsorption isotherms collected at 273 K (Extended Data Fig. 4a). RC-COF-1 showed a  $\text{CO}_2$  uptake of  $147 \text{ cm}^3 \text{ g}^{-1}$  (28.9 wt%) at 1 bar. This is to our knowledge the highest  $\text{CO}_2$  capacity reported in COFs under these measurement conditions<sup>44</sup> (Extended Data Fig. 4b). The calculated heat of adsorption was around  $35 \text{ kJ mol}^{-1}$  at the adsorption onset (Supplementary Fig. 33), which is comparable to related small-pore COFs with high  $\text{CO}_2$  uptakes<sup>44</sup>. The directly polymerized analogue, DP-COF-1, showed a similar heat of adsorption but a much lower  $\text{CO}_2$  uptake (Supplementary Fig. 32). RC-COF-1 also showed excellent chemical stability after treatment with concentrated HCl ( $12 \text{ mol l}^{-1}$ ) and NaOH ( $14 \text{ mol l}^{-1}$ ) solution for 24 h (Extended Data Fig. 5).

Scanning electron microscopy (SEM) showed that RC-COF-1 comprised uniform rod-like crystallites with an average size of around 600 nm (Fig. 3e), whereas DP-COF-1 was composed of less regular aggregates (Supplementary Fig. 35c). The high crystallinity for RC-COF-1 allowed us to confirm its periodic porous structure using high-resolution transmission electron microscopy (HRTEM). Reticular structures with hexagonal pores oriented perpendicular to the crystallographic *c* axis were observed (Fig. 3f). The calculated distance between the centres of two adjacent pores was 2.2 nm, in good agreement with the refined eclipsed model. Fast Fourier transform (FFT) conducted on a selected area showed a hexagonal symmetry; by contrast, no lattice fringes were discerned for DP-COF-1 prepared by direct polymerization (Supplementary Fig. 37a).

We next considered the generality of this reconstruction protocol for other COFs. For Urea-COF-2 (also known as COF-118; ref. 42), we used a commercially available isocyanate as the starting material; for Urea-COF-3 and Urea-COF-4, we used more widely accessible arylamine monomers, which could be easily converted into diureas before pre-organization (Fig. 1, Supplementary Fig. 7). Again, these three reconstructed COFs (RC-COF-2, RC-COF-3 and RC-COF-4) all showed superior crystallinity relative to COFs that were prepared by direct polymerization as per previously reported procedures (DP-COF-2 (also known as TpBD-Me<sub>2</sub>; ref. 45), DP-COF-3 (or TpBD; ref. 46) and DP-COF-4 (or TP-EDDA; ref. 24)), and all showed sharp and well-resolved diffraction peaks (Fig. 3g–i). Indeed, the difference in crystallinity levels between the pre-organized, reconstructed COF and its direct polycondensation analogue was even more pronounced for the mesoporous COF, RC-COF-4, which has the largest pores in this series of materials. Nitrogen adsorption measurements revealed greatly increased surface areas and pore volumes for the reconstructed COFs (Fig. 3j, Extended Data Fig. 2); BET surface areas increased from 623 m<sup>2</sup> g<sup>-1</sup> (DP-COF-2) to 2,792 m<sup>2</sup> g<sup>-1</sup> (RC-COF-2); from 573 m<sup>2</sup> g<sup>-1</sup> (DP-COF-3) to 2,461 m<sup>2</sup> g<sup>-1</sup> (RC-COF-3); and from 877 m<sup>2</sup> g<sup>-1</sup> (DP-COF-4) to 2,301 m<sup>2</sup> g<sup>-1</sup> (RC-COF-4). As such, the surface areas of the reconstructed COFs were between 2.6 and 4.5 times larger than the directly polymerized analogues. Likewise, the measured pore volumes were two to four times higher for the reconstructed COFs than for directly polymerized analogues. Pore size distribution profiles indicated the mesoporous nature of RC-COF-2, RC-COF-3 and RC-COF-4, with pore sizes of 2.3, 2.4 and 2.8 nm, respectively, in precise agreement with their structural models; whereas directly polymerized COFs show broader pore size distributions (Extended Data Fig. 3). HRTEM images (Fig. 3k,l, Supplementary Fig. 36) showed ordered porous structures extending through the crystal domains with clearly visible honeycomb pores, and the periodicities were consistent with the unit cell derived from the Pawley refined PXRD data. Few such ordered domains could be observed in the directly polymerized analogues (Supplementary Fig. 37). The reconstructed COFs also showed better thermal stability than the directly polymerized analogues, presumably because of their

enhanced crystallinity (Supplementary Fig. 34). The increased crystallinity in RC-COF-1 improves photogenerated charge carrier transport, leading to sacrificial photocatalytic hydrogen evolution rates of up to  $27.98 \text{ mmol h}^{-1} \text{ g}^{-1}$ . This is one of the highest activities reported for a COF photocatalyst and four times higher than the chemically equivalent but less ordered DP-COF-1 (Extended Data Figs. 6, 7, Supplementary Information).

## Density functional theory calculations

Density functional theory (DFT) calculations were used to investigate this reconstruction protocol in more detail. Direct polymerization of 1,3,5-triformylphloroglucinol with *p*-phenylenediamine yields rather low crystallinity because of the poorly reversible bond formation and the tautomerization into a stable  $\beta$ -ketoenamine form<sup>4</sup> (Fig. 4a, Supplementary Fig. 42a), which does not allow for full error correction. As a result, the directly polymerized product, DP-COF-1, does not attain the crystalline thermodynamic minimum structure (Fig. 4b). By contrast, decorating arylamine monomers with urea groups decreases the reactivity and enhances the reversibility for the bond formation (Supplementary Fig. 42b). This yields a highly crystalline but ‘soft’ urea precursor framework. Typically, hydrolysis of crystalline frameworks might lead to amorphization, but here, under appropriate solvothermal conditions (Fig. 2c), confinement in the framework coupled with fast imine condensation leads to the retention of crystallinity, together with a high conversion yield. The nature of the reconstruction process is revealed by the evolution of PXRD patterns recorded at different time intervals (Extended Data Fig. 1a). The diffraction peaks were found to shift continuously<sup>47</sup>, with no detectable disorder during the reconstruction. We suggest that the reconstruction in each COF crystallite has relatively slow kinetics with respect to the PXRD collection timescale. The gradual shift in peak positions indicates a smooth, continuous shrinkage of the lattice upon hydrolysis and re-polymerization. By contrast, if a rapid and concerted phase transformation was occurring, then we would expect to observe two distinct sets of interconverting PXRD peaks<sup>48</sup>, and no intermediate phases would be observed.

**Fig. 4: Reconstruction protocol with DFT calculations.**

---

 **figure 4**

**a, b**, Scheme showing the reaction paths for direct imine polycondensation and reconstruction synthesis (**a**). Direct polymerization yields only semi-crystalline COFs owing to the poor reversibility of the  $\beta$ -ketoenamine linkage, whereas decoration with urea groups decreases the reactivity and increases the reversibility, which leads to highly crystalline Urea-COFs (**b**); these then undergo framework reconstruction into  $\beta$ -ketoenamine RC-COFs with retained crystallinity. **c**, DFT-optimized geometries of phenylene amine molecules released by hydrolysis, confined on the surface of the reconstructing COF. The results of hydrolysis of a single urea bond (top) and both urea bonds (bottom) are shown. Coloured isosurfaces are the intermolecular interactions quantified by an independent gradient model (isosurface = 0.003 atomic units). In the bottom example, the interactions between the COF layers are omitted to highlight the interaction between the *p*-phenylenediamine monomer and the framework.

The simulation in Fig. [4c](#) suggests well-defined non-covalent interactions in the framework when the urea linkage is hydrolysed. When a single urea bond is cleaved (Fig. [4c](#)), strong hydrogen bonding is found in these simulations at the hydrolysis position between the resulting amine group and fragments of the COF in the same

layer, which maintains the position of the molecule in the framework. When both urea bonds are cleaved (Fig. 4c), the *p*-phenylenediamine molecule that is produced by hydrolysis is still captured by the framework through hydrogen bonds. The  $\pi$ - $\pi$  stacking between *p*-phenylenediamine and the adjacent COF layer reinforces these interactions (binding energy of around  $-16.08\text{ kcal mol}^{-1}$ ). We suggest that this nanoconfinement in the framework also stabilizes and protects the amine species by preventing the entry of other reactive species, contributing to the high yield of this multi-step reconstruction reaction. By contrast, such confinement is absent for the small-molecule model compounds (Fig. 1b) and the monomers can become disordered or react with other solution species once the urea linkages are hydrolysed.

## A convenient synthetic route

Vacuum sealing procedures are often necessary for COF syntheses to prevent oxidation of the arylamine monomers under solvothermal conditions. Freeze-pump-thaw procedures can be performed in a research laboratory, but they could represent a major hurdle for industrial scale-up. By decorating with urea groups, we decrease the reactivity of the monomers and provide better oxidation resistance, and hence this reconstruction approach can be conducted without any vacuum degassing steps (Extended Data Fig. 8, Supplementary Information), while affording equivalent crystallinity and porosity (BET surface area  $S_{\text{BET}} = 1,650\text{ m}^2\text{ g}^{-1}$  for RC-COF-1) relative to reactions that were performed with vacuum sealing ( $S_{\text{BET}} = 1,712\text{ m}^2\text{ g}^{-1}$ ). Such simple vacuum-free, aqueous processes might prove decisive for the viable commercial scale-up of highly crystalline COFs, for which vacuum degassing and inertization are costly.

## Outlook

COFs can be highly crystalline or physicochemically robust, but rarely both<sup>5</sup>. On the basis of an unexpected framework reconstruction in urea COFs, we have established a general and scalable vacuum-free protocol to synthesize highly crystalline imine frameworks by using a reversible and removable urea linkage as a disposable tether to pre-organize monomers before irreversible polymerization. This separates the crystallization process from the formation of robust framework bonds. The removable covalent tether is stronger and more directional compared to the non-covalent interactions in strategies in which monomers are pre-organized in molecular crystals. The superior level of structural order in reconstructed COFs presents new opportunities for applications such as gas adsorption and photocatalysis. We used urea tethers here, but it is likely that other covalent pre-organization chemistries could be devised to access other frameworks.

# Methods

## Chemicals

All reagents were obtained from Sigma-Aldrich, Fisher Chemical, Adamas, Jilin Chinese Academy of Sciences–Yanshen Technology or Shanghai Tensus Bio-Chem Technology and used as received. Carbon nitride was bought from Carbodeon. Solvents were obtained from commercial sources and used without further purification. A sulfone-decorated imine COF, FS-COF, was synthesized according to previous literature<sup>22</sup>.

## Liquid NMR spectroscopy

<sup>1</sup>H and <sup>13</sup>C NMR spectra were recorded in solution at 400 MHz and 100 MHz, respectively, using a Bruker Avance 400 NMR spectrometer.

## High-resolution mass spectrometry

The high-resolution mass spectrometry data were obtained using a Waters LCT Premier XE spectrometer.

## Powder X-ray diffraction

PXRD patterns were recorded on a Bruker D8 Advance diffractometer with Cu K $\alpha$  radiation with a voltage of 40 kV. Data were collected in the  $2\theta$  range of 2–40° with steps of 0.02°.

## Fourier transform infrared spectroscopy

The FTIR spectra were recorded on neat samples in the range of 4,000–650 cm<sup>-1</sup> on a PerkinElmer FTIR spectrometer equipped with a single reflection diamond ATR module.

## X-ray photoelectron spectroscopy

X-ray photoelectron spectroscopy (XPS) data were measured in powder form using an ESCALAB 250Xi instrument (Thermo Fisher Scientific) with a monochromatized Al K $\alpha$  line source.

## Elemental microanalyses

Elemental microanalyses were measured in the Research Center of Analysis and Test of East China University of Science and Technology using the EURO EA3000 Elemental Analyzer.

## Solid-state NMR spectroscopy

The solid-state  $^{13}\text{C}$  NMR spectra were recorded on a Bruker Avance 400 NMR spectrometer with CP-MAS at a  $^{13}\text{C}$  frequency of 100 MHz under 12 kHz spinning rate under MAS condition.

## Thermogravimetric analyses

Thermogravimetric analyses were performed on an EXSTAR6000 by heating samples at  $20\text{ }^{\circ}\text{C min}^{-1}$  under a nitrogen atmosphere to  $800\text{ }^{\circ}\text{C}$ .

## Gas adsorption analysis

Apparent surface areas were measured by nitrogen adsorption at 77.3 K using a Micromeritics ASAP 2020 volumetric adsorption analyser. Powder samples were degassed offline at 393 K for 12 h under a dynamic vacuum ( $10^{-5}$  bar). Before the adsorption test, the inert gas was removed using a high vacuum provided by the turbo molecular drag pump. The specific surface areas were evaluated using the BET model. Pore size distributions of COFs were obtained from fitting the nonlocal density functional theory to the adsorption data.

Low-pressure gas adsorption measurements of  $\text{CO}_2$  (273, 283, 293 and 308 K) were performed on MicrotacBELsorp Max and MaxII gas adsorption analysers. Ultrahigh-purity (higher than 99.999%)  $\text{CO}_2$  in compressed gas cylinders was used throughout all experiments. Samples were degassed at 393 K for 12 h before measurement.  $\text{CO}_2$  adsorption isotherms of each COF were then fitted with virial model equations as follows:

$$\begin{aligned} \$\$ \ln(p) = & \ln(N) + \frac{1}{T} \sum_{i=0}^m \{ \ln(i) \}^m \\ & \{ a \}_{i=0}^m \times N^m + \sum_{j=0}^n \{ \ln(j) \}^n \\ & \{ b \}_{j=0}^n \times N^n, \$\$ \end{aligned}$$

in which  $N$  is the amount adsorbed (or uptake) in  $\text{mmol g}^{-1}$ ;  $p$  is the pressure in kPa;  $T$  is the temperature in K; and  $m$  and  $n$  are multinomial coefficients that determine the isosteric heat.

The isosteric heat of each COF was calculated from the virial fitting adsorption isotherms by using the Clausius–Clapeyron equation, in which  $Q_{\text{st}}$  is the isosteric heat in  $\text{J mol}^{-1}$ ,  $T$  is the temperature in K,  $P$  is the pressure in kPa, and  $R$  is the gas constant ( $8.314 \text{ J K}^{-1} \text{ mol}^{-1}$ ):

$$-\{Q\}_{\{\rm st\}} = R \{T\}^2 \{(\frac{\partial \ln P}{\partial T})\}_{\{n\}}$$

## Scanning electron microscopy

COF morphologies were imaged using a field-emission scanning electron microscope (Helios G4 UC, Thermo Fisher Scientific).

## Transmission electron microscopy

Transmission electron microscopy (TEM) characterizations were performed on a Themis Z microscope (Thermo Fisher Scientific) equipped with two aberration correctors under 200 kV. To minimize the electron beam damage, a cryo-transfer TEM holder (Model 2550, Fischione Instruments) was used, and the temperature was set below  $-175^\circ\text{C}$  during TEM imaging.

## Ultraviolet-visible absorption spectroscopy

Ultraviolet (UV)-visible absorption spectra of the COFs were recorded on a PerkinElmer Lambda 950 UV-vis-NIR spectrometer by measuring the reflectance of powders in the solid state.

## Photoluminescence spectroscopy

Photoluminescence spectra were recorded on a Varian Cary Eclipse fluorescence spectrophotometer by measuring the powders in the solid state.

## Electron paramagnetic resonance spectroscopy

Electron paramagnetic resonance (EPR) spectra were acquired at room temperature under ambient conditions using a Bruker EMX-8/2.7 spectrometer. COF powders were taken in an EPR tube and excited with a 300-W Xe lamp using a 420-nm filter.

## Time-correlated single photon counting measurements

Time-correlated single photon counting measurements were performed on an Edinburgh Instruments LS980-D2S2-STM spectrometer equipped with picosecond-

pulsed LED excitation sources and an R928 detector, with a stop count rate below 3%. An EPL-375 diode ( $\lambda = 370.5$  nm, instrument response 100 ps, full width at half maximum, FWHM) with a 450-nm high-pass filter for emission detection was used. Suspensions were prepared by ultrasonication of the COF in water. The instrument response was measured with colloidal silica (LUDOX HS-40, Sigma-Aldrich) at the excitation wavelength without filter. Decay times were fitted in FAST software using suggested lifetime estimates.

## Photoelectrochemical measurements

Indium tin oxide (ITO) glasses were cleaned by sonication in ethanol and acetone for 30 min respectively, then dried under nitrogen flow. Two milligrams of COF was dispersed in 0.2 ml ethanol with  $\mu$ 110 ten Nafion solution (5 wt% in a mixture of lower aliphatic alcohols and water) and ultrasonicated for 20 min to give a homogenous suspension. ITO glass slides were covered with a copper mask giving an area of  $0.28\text{ cm}^2$ . Ten microlitres of the suspension was drop-casted on the ITO glass and dried overnight at room temperature. Electrochemical impedance spectroscopy and photocurrent response were performed using a Bio-Logic SP-200 electrochemical system. A three-electrode set-up was used with a working electrode (COF on ITO glass), counter electrode (platinum plate) and reference electrode (Ag/AgCl), and the bias voltage was  $-0.35$  V. A 300-W Newport Xe light source (model 6258, ozone-free) with a 420-nm filter was used to illuminate the samples. A solution of 0.5 M  $\text{Na}_2\text{SO}_4$  ( $\text{pH} = 6.8$ ) was used for measurement.

## Photocatalytic hydrogen evolution experiments

A quartz flask was charged with the photocatalyst powder (2.5 mg),  $0.1\text{ mol l}^{-1}$  ascorbic acid water solution (25 ml) and a certain amount of platinum (Pt) as a cocatalyst, using hexachloroplatinic acid as a Pt precursor. The resulting suspension was ultrasonicated until the photocatalyst was well-dispersed before degassing by  $\text{N}_2$  bubbling for 30 min. The reaction mixture was illuminated with a 300 W Newport Xe light source (model 6258, ozone-free) using appropriate filters for the time specified under atmospheric pressure. The Xe light source was cooled by water circulating through a metal jacket. The samples were first illuminated for 5 h to complete Pt photo-deposition; then the flask was degassed by  $\text{N}_2$  bubbling for 30 min followed by the photocatalysis reaction. Gas samples were taken with a gas-tight syringe and run on a Bruker 450-GC gas chromatograph. Hydrogen was detected with a thermal conductivity detector referencing against standard gas with a known concentration of hydrogen. Hydrogen dissolved in the reaction mixture was not measured and the pressure increase generated by the evolved hydrogen was not considered in the calculations. The rates were determined from a linear regression fit. After 5 h of

photocatalysis, no carbon monoxide associated with framework or ascorbic acid decomposition could be detected on a gas chromatography system equipped with a pulsed discharge detector.

For stability measurements, a flask was charged with 2.5 mg of COF photocatalyst, 0.1 mol l<sup>-1</sup> ascorbic acid water solution (25 ml) and a certain amount of Pt (3 wt%) as a cocatalyst, using hexachloroplatinic acid as a Pt precursor. The resulting suspension was ultrasonicated to obtain a well-dispersed suspension, then transferred into a quartz reactor connected to a closed gas system (Labsolar-6A, Beijing Perfectlight). The reaction mixture was evacuated several times to ensure complete removal of oxygen, and the pressure was set to 13.33 kPa. The reactor was irradiated in a 90° angle with a 300-W Xe light-source. The wavelength of the incident light was controlled using a 420-nm long-pass cut-off filter. The temperature of the reaction solution was maintained at 10 °C by circulation of cool water. The evolved gases were detected on an online gas chromatograph (Shimadzu GC 2014C) with a thermal conductive detector. After the photocatalysis experiment, the photocatalyst was recovered by washing with water then solvent exchange with methanol and tetrahydrofuran, respectively, before drying at 60 °C under a vacuum.

## Measurement of external quantum efficiencies

The external quantum efficiencies (EQEs) for the photocatalytic H<sub>2</sub> evolution were measured using monochromatic LED lamps ( $\lambda = 420, 490, 515$  and 595 nm, respectively). For the experiments, the photocatalyst (2.5 mg) with Pt loading was suspended in an aqueous solution containing ascorbic acid (0.1 mol l<sup>-1</sup>). The light intensity was measured with a ThorLabs S120VC photodiode power sensor controlled by a ThorLabs PM100D Power and Energy Meter Console. The EQEs were estimated using the equation:

$$\text{EQE} = \frac{2 \times N_u}{N_e} \times 100\%$$

where  $N_u$  is the number of photons absorbed by the photocatalyst and  $N_e$  is the number of electrons produced.

## Computational methods

Periodic DFT calculations were performed within the plane-wave pseudopotential formalism, using the Vienna ab initio simulation package (VASP) code<sup>49</sup>. The projector augmented-wave method was applied to describe the electron–ion

interactions<sup>50,51</sup>. A kinetic-energy cut-off of 500 eV was used to define the plane-wave basis set, and the electronic Brillouin zone was integrated using  $\Gamma$ -centred Monkhorst–Pack grids with the smallest allowed spacing between  $k$ -points (KSPACING) being 0.25 Å<sup>-1</sup>. Geometry optimizations were performed using the Perdew–Burke–Ernzerhof exchange–correlation functional with the DFT-D3(BJ) dispersion correction<sup>52,53,54</sup>. Tolerances of 10<sup>-6</sup> eV and 10<sup>-2</sup> eV Å<sup>-1</sup> were applied during the optimization of the Kohn–Sham wavefunctions and the geometry optimizations, respectively.

For crystal structures of COFs, both lattice parameters and atomic positions are allowed to change during geometry optimization. The electronic structures of the optimized RC-COF-1 and Urea-COF-1 structures were then computed using a screened hybrid exchange–correlation functional (HSE06), giving key electronic properties, such as band gap and electrostatic potential, for comparison of the COFs. Within periodic boundary conditions, the electronic eigenvalues are given with respect to an internal reference. To achieve valence band alignment, using a common vacuum level, so that band energies can be compared for the different COF structures, we followed an approach devised for determining the vacuum level of porous structures<sup>55</sup>.

For the binding model constructed for the hydrolysis products of the COF, the hydrolysis-released *p*-phenylenediamine monomer was assumed to be trapped in a three-layer COF model, with the first layer being decomposed. The periodic COF layer was parallel to the XY plane and separated from its periodic images along the Z direction by a vacuum of around 14 Å<sup>-1</sup>. The lattice parameters were fixed, and the atomic positions were fully optimized during this process.

The binding energy was computed using the following formula:

$$\$ \$ \varDelta E_{\text{bind}} = E_{\text{system}} - (E_{\text{monomer}} + E_{\text{framework}}), \$ \$$$

in which  $E_{\text{system}}$  is the energy of COF with the first layer hydrolysed,  $E_{\text{monomer}}$  and  $E_{\text{framework}}$  are the energies of *p*-phenylenediamine and framework, respectively, and the corresponding conformers were kept the same as in that system.

To visualize the intermolecular interactions between the *p*-phenylenediamine monomer and the COF fragment, we used the independent gradient model (IGM)<sup>56</sup>. The IGM method quantifies the net electron density gradient attenuation that is due to intermolecular interactions, identifying non-covalent interactions and generating data composed solely of intermolecular interactions for drawing the corresponding 3D isosurface representations. Here structures were extracted out from the periodic calculation result with no hydrogen atoms added to the fragment, because we used

pro-molecular level electron density here. The Multiwfn program<sup>57</sup> was used for IGM analyses and the VMD program<sup>58</sup> was used for visualization.

The geometries of the complexes were fully optimized by means of the hybrid M06-2X functional<sup>59</sup>. For all atoms, the def2-SVP basis set<sup>60,61</sup> was applied. No symmetry or geometry constraint was imposed during optimizations. The optimized geometries were verified as local minima on the potential energy surface by frequency computations at the same theoretical level. These calculations were performed with the Gaussian 16 suite of programs<sup>62</sup>. Water was used as the solvent in the SMD solvation model<sup>63</sup>. A temperature of 433 K was used for thermochemistry analysis in all calculations.

## Data availability

The experimental and theoretical data that support the findings of this study are available from the corresponding authors upon reasonable request. [Source data](#) are provided with this paper.

## References

1. Côté, A. P. et al. Porous, crystalline, covalent organic frameworks. *Science* **310**, 1166–1170 (2005).
2. Diercks, C. S. & Yaghi, O. M. The atom, the molecule, and the covalent organic framework. *Science* **355**, eaal1585 (2017).
3. Slater, A. G. & Cooper, A. I. Function-led design of new porous materials. *Science* **348**, aaa8075 (2015).
4. Kandambeth, S. et al. Construction of crystalline 2D covalent organic frameworks with remarkable chemical (acid/base) stability via a combined reversible and irreversible route. *J. Am. Chem. Soc.* **134**, 19524–19527 (2012).
5. Haase, F. & Lotsch, B. V. Solving the COF trilemma: towards crystalline, stable and functional covalent organic frameworks. *Chem. Soc. Rev.* **49**, 8469–8500 (2020).
6. Beaudoin, D., Maris, T. & Wuest, J. D. Constructing monocrystalline covalent organic networks by polymerization. *Nat. Chem.* **5**, 830–834 (2013).
7. Ma, T. et al. Single-crystal x-ray diffraction structures of covalent organic frameworks. *Science* **361**, 48–52 (2018).

8. Evans, A. M. et al. Seeded growth of single-crystal two-dimensional covalent organic frameworks. *Science* **361**, 52–57 (2018).
9. Hu, Y. et al. Single crystals of mechanically entwined helical covalent polymers. *Nat. Chem.* **13**, 660–665 (2021).
10. Ascherl, L. et al. Molecular docking sites designed for the generation of highly crystalline covalent organic frameworks. *Nat. Chem.* **8**, 310–316 (2016).
11. Guan, X. et al. Chemically stable polyarylether-based covalent organic frameworks. *Nat. Chem.* **11**, 587–594 (2019).
12. Jin, E. et al. Two-dimensional sp<sup>2</sup> carbon-conjugated covalent organic frameworks. *Science* **357**, 673–676 (2017).
13. Li, X. et al. Partitioning the interlayer space of covalent organic frameworks by embedding pseudorotaxanes in their backbones. *Nat. Chem.* **12**, 1115–1122 (2020).
14. Zhao, S. et al. Hydrophilicity gradient in covalent organic frameworks for membrane distillation. *Nat. Mater.* **20**, 1551–1558 (2021).
15. Liang, R. et al. Rational design of crystalline two-dimensional frameworks with highly complicated topological structures. *Nat. Commun.* **10**, 4609 (2019).
16. Zhang, L. et al. Covalent organic framework for efficient two-photon absorption. *Matter* **2**, 1049–1063 (2020).
17. Wang, Z., Zhang, S., Chen, Y., Zhang, Z. & Ma, S. Covalent organic frameworks for separation applications. *Chem. Soc. Rev.* **49**, 708–735 (2020).
18. Dong, J., Han, X., Liu, Y., Li, H. & Cui, Y. Metal–covalent organic frameworks (MCOFs): a bridge between metal–organic frameworks and covalent organic frameworks. *Angew. Chem. Int. Edn* **59**, 13722–13733 (2020).
19. Zhu, H. et al. Efficient electron transmission in covalent organic framework nanosheets for highly active electrocatalytic carbon dioxide reduction. *Nat. Commun.* **11**, 497 (2020).
20. Banerjee, T., Podjaski, F., Kröger, J., Biswal, B. P. & Lotsch, B. V. Polymer photocatalysts for solar-to-chemical energy conversion. *Nat. Rev. Mater.* **6**, 168–190 (2021).

21. Vyas, V. S. et al. A tunable azine covalent organic framework platform for visible light-induced hydrogen generation. *Nat. Commun.* **6**, 8508 (2015).
22. Wang, X. et al. Sulfone-containing covalent organic frameworks for photocatalytic hydrogen evolution from water. *Nat. Chem.* **10**, 1180–1189 (2018).
23. Wang, Y. et al. Current understanding and challenges of solar-driven hydrogen generation using polymeric photocatalysts. *Nat. Energy* **4**, 746–760 (2019).
24. Pachfule, P. et al. Diacetylene functionalized covalent organic framework (COF) for photocatalytic hydrogen generation. *J. Am. Chem. Soc.* **140**, 1423–1427 (2018).
25. Jin, E. et al. 2D sp<sup>2</sup> carbon-conjugated covalent organic frameworks for photocatalytic hydrogen production from water. *Chem* **5**, 1632–1647 (2019).
26. Ghosh, S. et al. Identification of prime factors to maximize the photocatalytic hydrogen evolution of covalent organic frameworks. *J. Am. Chem. Soc.* **142**, 9752–9762 (2020).
27. Bi, S. et al. Two-dimensional semiconducting covalent organic frameworks via condensation at arylmethyl carbon atoms. *Nat. Commun.* **10**, 2467 (2019).
28. Chen, R. et al. Rational design of isostructural 2D porphyrin-based covalent organic frameworks for tunable photocatalytic hydrogen evolution. *Nat. Commun.* **12**, 1354 (2021).
29. Chen, W. et al. Modulating benzothiadiazole-based covalent organic frameworks via halogenation for enhanced photocatalytic water splitting. *Angew. Chem. Int. Edn* **59**, 16902–16909 (2020).
30. Colson, J. W. et al. Oriented 2D covalent organic framework thin films on single-layer graphene. *Science* **332**, 228–231 (2011).
31. Wang, M. et al. Unveiling electronic properties in metal-phthalocyanine-based pyrazine-linked conjugated two-dimensional covalent organic frameworks. *J. Am. Chem. Soc.* **141**, 16810–16816 (2019).
32. Kandambeth, S., Dey, K. & Banerjee, R. Covalent organic frameworks: chemistry beyond the structure. *J. Am. Chem. Soc.* **141**, 1807–1822 (2019).
33. Li, X. et al. Chemically robust covalent organic frameworks: progress and perspective. *Matter* **3**, 1507–1540 (2020).

34. Kissel, P. et al. A two-dimensional polymer prepared by organic synthesis. *Nat. Chem.* **4**, 287–291 (2012).
35. Kory, M. J. et al. Gram-scale synthesis of two-dimensional polymer crystals and their structure analysis by x-ray diffraction. *Nat. Chem.* **6**, 779–784 (2014).
36. Kissel, P., Murray, D. J., Wulf Lange, W. J., Catalano, V. J. & King, B. T. A nanoporous two-dimensional polymer by single-crystal-to-single-crystal photopolymerization. *Nat. Chem.* **6**, 774–778 (2014).
37. Liu, W. et al. A two-dimensional conjugated aromatic polymer via C–C coupling reaction. *Nat. Chem.* **9**, 563–570 (2017).
38. Liu, W. et al. Two-dimensional polymer synthesized via solid-state polymerization for high-performance supercapacitors. *ACS Nano* **12**, 852–860 (2018).
39. Liu, K. et al. On-water surface synthesis of crystalline, few-layer two-dimensional polymers assisted by surfactant monolayers. *Nat. Chem.* **11**, 994–1000 (2019).
40. Grommet, A. B., Feller, M. & Klajn, R. Chemical reactivity under nanoconfinement. *Nat. Nanotechnol.* **15**, 256–271 (2020).
41. Iwasawa, T., Hooley, R. J. & Rebek, J. Stabilization of labile carbonyl addition intermediates by a synthetic receptor. *Science* **317**, 493–496 (2007).
42. Zhao, C. et al. Urea-linked covalent organic frameworks. *J. Am. Chem. Soc.* **140**, 16438–16441 (2018).
43. Ma, T. et al. Diverse crystal size effects in covalent organic frameworks. *Nat. Commun.* **11**, 6128 (2020).
44. Gao, Q. et al. Covalent organic framework with frustrated bonding network for enhanced carbon dioxide storage. *Chem. Mater.* **30**, 1762–1768 (2018).
45. Chandra, S. et al. Chemically stable multilayered covalent organic nanosheets from covalent organic frameworks via mechanical delamination. *J. Am. Chem. Soc.* **135**, 17853–17861 (2013).
46. Biswal, B. P. et al. Mechanochemical synthesis of chemically stable isoreticular covalent organic frameworks. *J. Am. Chem. Soc.* **135**, 5328–5331 (2013).

47. Carrington, E. J. et al. Solvent-switchable continuous-breathing behaviour in a diamondoid metal–organic framework and its influence on CO<sub>2</sub> versus CH<sub>4</sub> selectivity. *Nat. Chem.* **9**, 882–889 (2017).
48. Reed, D. A. et al. A spin transition mechanism for cooperative adsorption in metal–organic frameworks. *Nature* **550**, 96–100 (2017).
49. Kresse, G. & Furthmüller, J. Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set. *Phys. Rev. B* **54**, 11169–11186 (1996).
50. Kresse, G. & Joubert, D. From ultrasoft pseudopotentials to the projector augmented-wave method. *Phys. Rev. B* **59**, 1758 (1999).
51. Blöchl, P. E. Projector augmented-wave method. *Phys. Rev. B* **50**, 17953 (1994).
52. Perdew, J. P., Burke, K. & Ernzerhof, M. Generalized gradient approximation made simple. *Phys. Rev. Lett.* **77**, 3865 (1996).
53. Grimme, S., Antony, J., Ehrlich, S. & Krieg, H. A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H–Pu. *J. Chem. Phys.* **132**, 154104 (2010).
54. Grimme, S., Ehrlich, S. & Goerigk, L. Effect of the damping function in dispersion corrected density functional theory. *J. Comput. Chem.* **32**, 1456–1465 (2011).
55. Butler, K. T., Hendon, C. H. & Walsh, A. Electronic chemical potentials of porous metal–organic frameworks. *J. Am. Chem. Soc.* **136**, 2703–2706 (2014).
56. Lefebvre, C. et al. Accurately extracting the signature of intermolecular interactions present in the NCI plot of the reduced density gradient versus electron density. *Phys. Chem. Chem. Phys.* **19**, 17928–17936 (2017).
57. Lu, T. & Chen, F. Multiwfn: a multifunctional wavefunction analyzer. *J. Comput. Chem.* **33**, 580–592 (2012).
58. Humphrey, W., Dalke, A. & Schulten, K. VMD: visual molecular dynamics. *J. Mol. Graph.* **14**, 33–38 (1996).
59. Zhao, Y. & Truhlar, D. G. The M06 suite of density functionals for main group thermochemistry, thermochemical kinetics, noncovalent interactions, excited states, and transition elements: two new functionals and systematic testing of four M06-class functionals and 12 other functionals. *Theor. Chem. Acc.* **120**, 215–241 (2008).

60. Weigend, F. & Ahlrichs, R. Balanced basis sets of split valence, triple zeta valence and quadruple zeta valence quality for H to Rn: design and assessment of accuracy. *Phys. Chem. Chem. Phys.* **7**, 3297–3305 (2005).
61. Weigend, F. Accurate Coulomb-fitting basis sets for H to Rn. *Phys. Chem. Chem. Phys.* **8**, 1057–1065 (2006).
62. Frisch, M. et al. Gaussian 16, Revision A.03 (Gaussian Inc., 2016).
63. Marenich, A. V., Cramer, C. J. & Truhlar, D. G. Universal solvation model based on solute electron density and on a continuum model of the solvent defined by the bulk dielectric constant and atomic surface tensions. *J. Phys. Chem. B* **113**, 6378–6396 (2009).

## Acknowledgements

We gratefully acknowledge funding from the National Natural Science Foundation of China (NSFC) Science Center Program (21788102); the NSFC (21905091, 21636002); Shanghai Municipal Science and Technology Major Project (2018SHZDZX03, 21JC1401700); Shanghai Municipal Science and Technology (20120710200); the Engineering and Physical Sciences Research Council (EPSRC) (EP/N004884/1, EP/P034497/1, EP/S017623/1); the Leverhulme Trust through the Leverhulme Research Centre for Functional Materials Design; and China Postdoctoral Science Foundation (no. 2019M651418). S.D. acknowledges support from the Shanghai Rising Star Program (20QA1402400) and the Program for Professor of Special Appointment (Eastern Scholar) at Shanghai Institutions of Higher Learning. W.Z. thanks the Research Center of Analysis and Testing of East China University of Science and Technology for the characterization of elemental microanalyses.

## Author information

### Affiliations

1. Key Laboratory for Advanced Materials and Institute of Fine Chemicals, Joint International Research Laboratory of Precision Chemistry and Molecular Engineering, Feringa Nobel Prize Scientist Joint Research Center, Frontiers Science Center for Materiobiology and Dynamic Chemistry, School of Chemistry and Molecular Engineering, East China University of Science and Technology, Shanghai, China

Weiwei Zhang, Linjiang Chen, Sheng Dai, Chengxi Zhao, Miaojie Yu, Shuhao An, Xiaofeng Wu, Yongzhen Wu, He Tian, Wei-Hong Zhu & Andrew I. Cooper

2. Leverhulme Centre for Functional Materials Design, Materials Innovation Factory and Department of Chemistry, University of Liverpool, Liverpool, UK

Linjiang Chen, Chengxi Zhao, Samantha Y. Chong, Haofan Yang, Lunjie Liu, Yang Bai, Yongjie Xu, Xiao-Wei Zhu, Qiang Zhu, Reiner Sebastian Sprick, Marc A. Little, Xiaofeng Wu & Andrew I. Cooper

3. State Key Laboratory of Chemical Engineering, East China University of Science and Technology, Shanghai, China

Cheng Ma & Minghui Zhu

4. School of Physical Science and Technology, ShanghaiTech University, Shanghai, China

Lei Wei, Shan Jiang & Yue-Biao Zhang

## Contributions

W.Z., H.T., W.-H.Z. and A.I.C. conceived the project. W.Z. and M.Y. synthesized and characterized the materials, and performed photocatalysis experiments. L.C. and C.Z. performed the simulations. S.D. performed SEM and HRTEM characterizations. M.Z. and Q.Z. conducted FTIR characterizations. L.C., L.W., Y.-B.Z. and S.Y.C. conducted structural modelling and PXRD refinements. S.J. performed solid-state NMR measurement. C.M., Y.-B.Z. and S.A. conducted the gas adsorption experiments. H.Y., Y.B. and X.-W.Z. carried out optical spectra measurements. L.L. performed photo-electrochemistry characterizations. Y.X. conducted XPS experiments. W.Z., S.D., R.S.S., M.A.L., X.W., Y.W., H.T., W.-H.Z. and A.I.C. analysed the data and wrote the paper. All authors discussed the results and contributed to the manuscript.

## Corresponding authors

Correspondence to [Wei-Hong Zhu](#) or [Andrew I. Cooper](#).

## Ethics declarations

### Competing interests

The authors declare no competing interests.

## Peer review

## Peer review information

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

## Additional information

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## Extended data figures and tables

### Extended Data Fig. 1 Continuous transformation in water.

**a, b,** The evolution of the PXRD patterns with time for as-synthesized Urea-COF-1 by treatment with water at the reaction temperatures of 160 °C (samples were solvated with water (**a**) and measured after activation/desolvation (**b**)). The gradual shift in the peak positions over time suggests a continuous transformation during the COF reconstruction.

[Source data](#)

### Extended Data Fig. 2 Increased porosity.

**a, b,** Comparison of BET surface areas (**a**) and total pore volumes (**b**) for RC-COFs (yellow bars) and DP-COFs (green bars). Reconstructed COFs showed greatly improved BET surface area and pore volumes than directly polymerized analogues.

### Extended Data Fig. 3 Pore size distributions.

**a–h,** Pore size distribution profiles for RC-COFs (**a, c, e, g**) and DP-COFs (**b, d, f, h**). The reconstructed COFs showed a much narrower pore size distribution, which reflected their enhanced crystallinity.

[Source data](#)

### Extended Data Fig. 4 CO<sub>2</sub> uptake.

**a, b,** Adsorption (filled circles) and desorption (open circles) isotherms for CO<sub>2</sub> uptake in RC-COF-1 and DP-COF-1 recorded at 273 K (**a**). Summary of CO<sub>2</sub> uptake

capacities for COF materials reported at 273 K and 1 bar (**b**); see details in Supplementary Table [5](#).

[Source data](#)

**Extended Data Fig. 5 Chemical stability.**

**a, b**, PXRD patterns (**a**) and FTIR spectra (**b**) for RC-COF-1 after various chemical treatments for 24 h. RC-COF-1 retained its crystalline structure after treatment under all these conditions. Note that the PXRD peak intensity decreased slightly when treated with concentrated sodium hydroxide solution.

[Source data](#)

**Extended Data Fig. 6 Optical and electronic properties and photocatalytic hydrogen evolution activity.**

**a**, EPR studies showed a single Lorentzian line centred at a *g* value of 2.006 for RC-COF-1, which intensified dramatically upon light excitation, suggesting an effective light-induced charge carrier generation, whereas DP-COF-1 displayed much lower signal intensity under same test conditions. **b**, Transient photocurrents with on-off light intermittent irradiation ( $\lambda > 420$  nm) for RC-COF-1, DP-COF-1 and Urea-COF-1, conducted with a bias potential of  $-0.35$  V vs Ag/AgCl. RC-COF-1 produced an enhanced photocurrent compared to its semi-crystalline counterpart, DP-COF-1, indicating more efficient separation of photogenerated charge carriers. **c**, Time courses of sacrificial photocatalytic hydrogen production for RC-COF-1, FS-COF, DP-COF-1, Urea-COF-1 and g-C<sub>3</sub>N<sub>4</sub> (2.5 mg catalyst in water with 3 wt.% Pt loading,  $\lambda > 420$  nm for RC-COF-1, FS-COF, DP-COF-1 and Urea-COF-1, and  $\lambda > 295$  nm for g-C<sub>3</sub>N<sub>4</sub>). **d**, Time course of photocatalytic hydrogen evolution for RC-COF-1 from three different synthetic batches under visible light irradiation; inset is corresponding hydrogen evolution rate (HER). There is good batch-to-batch reproducibility in terms of photocatalytic performance for materials prepared by this reconstruction route. **e**, The external quantum efficiencies (EQEs) of RC-COF-1 were estimated to be 6.39% at 420 nm, 5.92% at 490 nm, 5.20% at 515 nm, and 1.62% at 595 nm, respectively. By comparison, DP-COF-1 exhibited a much-lower EQEs of 1.97%, 1.61%, 1.37%, and 0.54% at the same wavelengths. **f**, Long-term photocatalytic hydrogen evolution stability test for RC-COF-1 over 60 h under visible light ( $\lambda > 420$  nm). The dashed vertical lines denote degassing and addition of a further 1.25 mmol of ascorbic acid. No obvious decrease in activity was observed during this 60-h period.

[Source data](#)

## **Extended Data Fig. 7 RC-COF-1 crystallite with photo-deposited Pt co-catalyst nanoparticles.**

**a–g**, SEM (**a**), TEM (**b**), high angle annular dark field scanning transmission electron microscopy (HAADF-STEM) (**c**) images and elemental mapping (**d–g**) for RC-COF-1 crystallite decorated with photo-deposited Pt co-catalyst. The inset in **b** shows uniform distributions of Pt nanoparticles ( $2.5 \pm 0.5$  nm) in the selected area (yellow square). The uniform morphology of the reconstructed COF and the good Pt cocatalyst dispersion might also contribute to its enhanced activity.

[Source data](#)

## **Extended Data Fig. 8 RC-COF-1 synthesized without vacuum degassing step.**

**a, b**, Experimental PXRD patterns (**a**) and nitrogen adsorption isotherm (**b**, filled symbols) and desorption isotherm (**b**, open symbols) for RC-COF-1 synthesized without vacuum degassing steps; inset shows calculated pore size distribution. Both the crystallinity and the porous properties are close to those obtained with careful degassing, suggesting that this step is unnecessary in the reconstruction synthesis route, at least for these specific monomers.

[Source data](#)

## **Supplementary information**

### **Supplementary Information**

This file contains supplementary text, supplementary tables 1–5, supplementary figures 1–76 and supplementary references.

### **Supplementary Data**

This file contains coordinates files.

## **Source data**

### **Source Data Fig. 2**

### **Source Data Fig. 3**

[\*\*Source Data Extended Data Fig. 1\*\*](#)

[\*\*Source Data Extended Data Fig. 3\*\*](#)

[\*\*Source Data Extended Data Fig. 4\*\*](#)

[\*\*Source Data Extended Data Fig. 5\*\*](#)

[\*\*Source Data Extended Data Fig. 6\*\*](#)

[\*\*Source Data Extended Data Fig. 7\*\*](#)

[\*\*Source Data Extended Data Fig. 8\*\*](#)

## 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

Zhang, W., Chen, L., Dai, S. *et al.* Reconstructed covalent organic frameworks. *Nature* **604**, 72–79 (2022). <https://doi.org/10.1038/s41586-022-04443-4>

- Received: 29 July 2021
- Accepted: 18 January 2022
- Published: 06 April 2022

- Issue Date: 07 April 2022
- DOI: <https://doi.org/10.1038/s41586-022-04443-4>

## Share this article

Anyone you share the following link with will be able to read this content:

[Get shareable link](#)

Sorry, a shareable link is not currently available for this article.

[Copy to clipboard](#)

Provided by the Springer Nature SharedIt content-sharing initiative

---

This article was downloaded by **calibre** from <https://www.nature.com/articles/s41586-022-04443-4>

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

- Article
- [Published: 06 April 2022](#)

# Autonomous fuelled directional rotation about a covalent single bond

- [Stefan Borsley](#) [ORCID: orcid.org/0000-0001-9637-4178](#)<sup>1 na1</sup>,
- [Elisabeth Kreidt](#) [ORCID: orcid.org/0000-0003-1229-3181](#)<sup>1 na1</sup>,
- [David A. Leigh](#) [ORCID: orcid.org/0000-0002-1202-4507](#)<sup>1,2</sup> &
- [Benjamin M. W. Roberts](#) [ORCID: orcid.org/0000-0003-2820-8359](#)<sup>1 na1</sup>

[Nature](#) volume **604**, pages 80–85 (2022)

- 3532 Accesses
- 199 Altmetric
- [Metrics details](#)

## Subjects

- [Molecular machines and motors](#)
- [Supramolecular chemistry](#)

## Abstract

Biology operates through autonomous chemically fuelled molecular machinery<sup>1</sup>, including rotary motors such as adenosine triphosphate synthase<sup>2</sup> and the bacterial flagellar motor<sup>3</sup>. Chemists have long sought to

create analogous molecular structures with chemically powered, directionally rotating, components<sup>4,5,6,7,8,9,10,11,12,13,14,15,16,17</sup>. However, synthetic motor molecules capable of autonomous 360° directional rotation about a single bond have proved elusive, with previous designs lacking either autonomous fuelling<sup>7,10,12</sup> or directionality<sup>6</sup>. Here we show that 1-phenylpyrrole 2,2'-dicarboxylic acid<sup>18,19</sup> (**1a**) is a catalysis-driven<sup>20,21</sup> motor that can continuously transduce energy from a chemical fuel<sup>9,20,21,22,23,24,25,26,27</sup> to induce repetitive 360° directional rotation of the two aromatic rings around the covalent N–C bond that connects them. On treatment of **1a** with a carbodiimide<sup>21,25,26,27</sup>, intramolecular anhydride formation between the rings and the anhydride's hydrolysis both occur incessantly. Both reactions are kinetically gated<sup>28,29,30</sup> causing directional bias. Accordingly, catalysis of carbodiimide hydration by the motor molecule continuously drives net directional rotation around the N–C bond. The directionality is determined by the handedness of both an additive that accelerates anhydride hydrolysis and that of the fuel, and is easily reversed additive<sup>31</sup>. More than 97% of fuel molecules are consumed through the chemical engine cycle<sup>24</sup> with a directional bias of up to 71:29 with a chirality-matched fuel and additive. In other words, the motor makes a ‘mistake’ in direction every three to four turns. The 26-atom motor molecule's simplicity augurs well for its structural optimization and the development of derivatives that can be interfaced with other components for the performance of work and tasks<sup>32,33,34,35,36</sup>.

This is a preview of subscription content

## Access options

Subscribe to Nature+

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

\$29.99

monthly

## Subscribe

Subscribe to Journal

Get full journal access for 1 year

\$199.00

only \$3.90 per issue

## Subscribe

All prices are NET prices.

VAT will be added later in the checkout.

Tax calculation will be finalised during checkout.

Buy article

Get time limited or full article access on ReadCube.

\$32.00

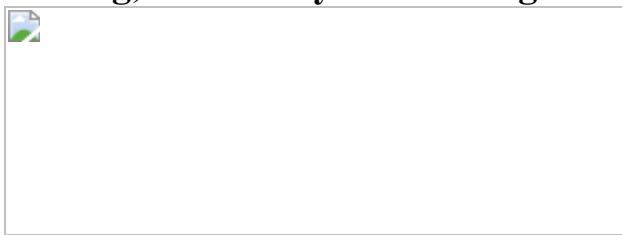
## Buy

All prices are NET prices.

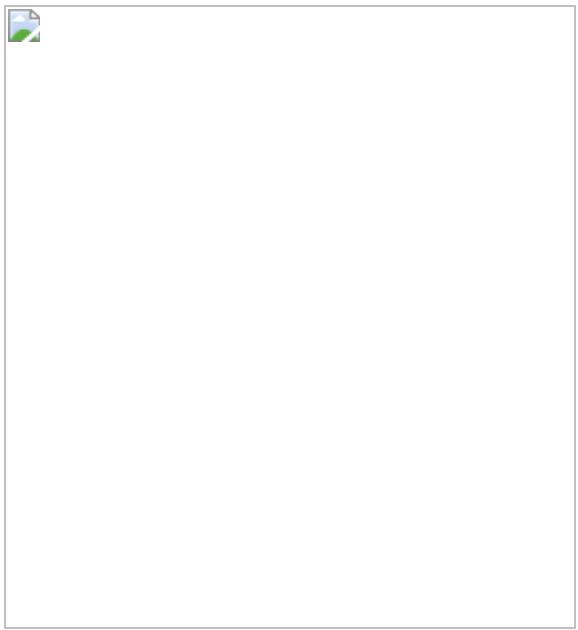
## **Additional access options:**

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

**Fig. 1: Chemical engine cycle of an autonomous, continuously operating, chemically fuelled single bond rotary motor.**



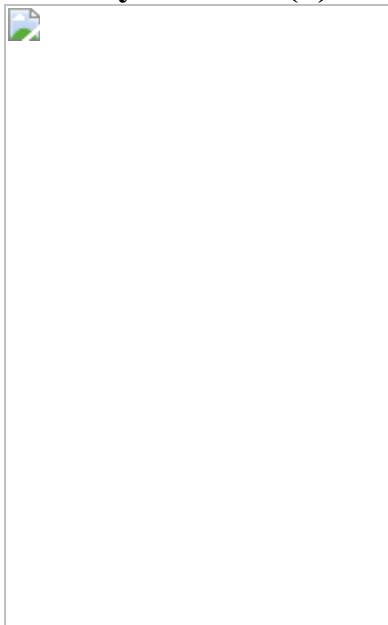
**Fig. 2: Mechanical gating of 1-arylpyrrole 2,2'-dicarboxylic acids (1).**



**Fig. 3: Chemical transformations of 1-arylpvrrole 2,2'-dicarboxylic acids (1).**



**Fig. 4: Autonomous chemically fuelled operation of 1-arylpvrrole 2,2'-dicarboxylic acids (1).**



## Data availability

The data that support the findings of this study are available within the paper and its [Supplementary Information](#), or are available from the Mendeley data repository (<https://data.mendeley.com/>) at <http://doi.org/10.17632/2zvst4kg2w.1>.

## References

1. Schliwa, M. & Woehlke, G. Molecular motors. *Nature* **422**, 759–765 (2003).
2. Boyer, P. D. Energy, life, and ATP (Nobel lecture). *Angew. Chem. Int. Ed.* **37**, 2296–2307 (1998).
3. Santiveri, M. et al. Structure and function of stator units of the bacterial flagellar motor. *Cell* **183**, 244–257 (2020).
4. Kelly, T. R., Tellitu, I. & Sestelo, J. P. In search of molecular ratchets. *Angew. Chem. Int. Ed. Engl.* **36**, 1866–1868 (1997).
5. Kelly, T. R., De Silva, H. & Silva, R. A. Unidirectional rotary motion in a molecular system. *Nature* **401**, 150–152 (1999).
6. Mock, W. L. & Ochwat, K. J. Theory and example of a small-molecule motor. *J. Phys. Org. Chem.* **16**, 175–182 (2003).
7. Fletcher, S. P., Dumur, F., Pollard, M. M. & Feringa, B. L. A reversible, unidirectional molecular rotary motor driven by chemical energy. *Science* **310**, 80–82 (2005).
8. Dahl, B. J. & Branchaud, B. P. 180° unidirectional bond rotation in a biaryl lactone artificial molecular motor prototype. *Org. Lett.* **8**, 5841–5844 (2006).
9. Wilson, M. R. et al. An autonomous chemically fuelled small-molecule motor. *Nature* **534**, 235–240 (2016).

10. Collins, B. S. L., Kistemaker, J. C. M., Otten, E. & Feringa, B. L. A chemically powered unidirectional rotary molecular motor based on a palladium redox cycle. *Nat. Chem.* **8**, 860–866 (2016).
11. Erbas-Cakmak, S. et al. Rotary and linear molecular motors driven by pulses of a chemical fuel. *Science* **358**, 340–343 (2017).
12. Zhang, Y. et al. A chemically driven rotary molecular motor based on reversible lactone formation with perfect unidirectionality. *Chem* **6**, 2420–2429 (2020).
13. Koumura, N., Zijlstra, R. W. J., van Delden, R. A., Harada, N. & Feringa, B. L. Light-driven monodirectional molecular rotor. *Nature* **401**, 152–155 (1999).
14. Pooler, D. R. S., Lubbe, A. S., Crespi, S. & Feringa, B. L. Designing light-driven rotary molecular motors. *Chem. Sci.* **12**, 14964–14986 (2021).
15. Kelly, T. R. et al. Progress toward a rationally designed, chemically powered rotary molecular motor. *J. Am. Chem. Soc.* **129**, 376–386 (2007).
16. Feynman, R. P., Leighton, R. B. & Sands, M. *The Feynman Lectures on Physics* Vol. 1, Ch. 46 (Addison-Wesley Publishing Company, 1963).
17. Davis, A. P. Tilting at windmills? The second law survives. *Angew. Chem. Int. Ed.* **37**, 909–910 (1998).
18. Fogassy, K. et al. Efficient synthesis and resolution of ( $\pm$ )-1-[2-carboxy-6-(trifluoromethyl)phenyl]pyrrole-2-carboxylic acid. *Tetrahedron Asymmetry* **11**, 4771–4780 (2000).
19. Faigl, F., Tárkányi, G., Fogassy, K., Tepfenhardt, D. & Thurner, A. Synthesis and stereochemical stability of new atropisomeric 1-(substituted phenyl)pyrrole derivatives. *Tetrahedron* **64**, 1371–1377 (2008).

20. Amano, S., Fielden, S. D. P. & Leigh, D. A. A catalysis-driven artificial molecular pump. *Nature* **594**, 529–534 (2021).
21. Borsley, S., Leigh, D. A. & Roberts, B. M. W. A doubly kinetically-gated information ratchet autonomously driven by carbodiimide hydration. *J. Am. Chem. Soc.* **143**, 4414–4420 (2021).
22. Ragazzon, G. & Prins, L. J. Energy consumption in chemical fuel-driven self-assembly. *Nat. Nanotechnol.* **13**, 882–889 (2018).
23. Kariyawasam, L. S., Hossain, M. M. & Hartley, C. S. The transient covalent bond in abiotic nonequilibrium systems. *Angew. Chem. Int. Ed.* **60**, 12648–12658 (2021).
24. Amano, S., Borsley, S., Leigh, D. A. & Sun, Z. Chemical engines: driving systems away from equilibrium through catalyst reaction cycles. *Nat. Nanotechnol.* **16**, 1057–1067 (2021).
25. Tena-Solsona, M. et al. Non-equilibrium dissipative supramolecular materials with a tunable lifetime. *Nat. Commun.* **8**, 15895 (2017).
26. Kariyawasam, L. S. & Hartley, C. S. Dissipative assembly of aqueous carboxylic acid anhydrides fueled by carbodiimides. *J. Am. Chem. Soc.* **139**, 11949–11955 (2017).
27. Bal, S., Das, K., Ahmed, S. & Das, D. Chemically fueled dissipative self-assembly that exploits cooperative catalysis. *Angew. Chem. Int. Ed.* **58**, 244–247 (2019).
28. Astumian, R. D. Irrelevance of the power stroke for the directionality, stopping force, and optimal efficiency of chemically driven molecular machines. *Biophys. J.* **108**, 291–303 (2015).
29. Erbas-Cakmak, S., Leigh, D. A., McTernan, C. T. & Nussbaumer, A. L. Artificial molecular machines. *Chem. Rev.* **115**, 10081–10206 (2015).

30. Astumian, R. D., Mukherjee, S. & Warshel, A. The physics and physical chemistry of molecular machines. *ChemPhysChem* **17**, 1719–1741 (2016).
31. Dálaigh, C. Ó. & Connon, S. J. Nonenzymatic acylative kinetic resolution of Baylis-Hillman adducts. *J. Org. Chem.* **72**, 7066–7069 (2007).
32. Li, Q. et al. Macroscopic contraction of a gel induced by the integrated motion of light-driven molecular motors. *Nat. Nanotechnol.* **10**, 161–165 (2015).
33. García-López, V. et al. Molecular machines open cell membranes. *Nature* **548**, 567–572 (2017).
34. Feng, L. et al. Active mechanisorption driven by pumping cassettes. *Science* **374**, 1215–1221 (2021).
35. Thomas, D., et al. Pumping between phases with a pulsed-fuel molecular ratchet. Preprint at <https://doi.org/10.33774/chemrxiv-2021-fl7tv> (2021).
36. Zhang, Q. et al. Muscle-like artificial molecular actuators for nanoparticles. *Chem* **4**, 2670–2684 (2018).
37. Astumian, R. D. & Bier, M. Mechanochemical coupling of the motion of molecular motors to ATP hydrolysis. *Biophys. J.* **70**, 637–653 (1996).
38. Astumian, R. D. Thermodynamics and kinetics of a Brownian Motor. *Science* **276**, 917–922 (1997).
39. Serreli, V., Lee, C.-F., Kay, E. R. & Leigh, D. A. A molecular information ratchet. *Nature* **445**, 523–527 (2007).
40. Alvarez-Pérez, M., Goldup, S. M., Leigh, D. A. & Slawin, A. M. Z. A chemically-driven molecular information ratchet. *J. Am. Chem. Soc.* **130**, 1836–1838 (2008).

41. Astumian, R. D. Kinetic asymmetry allows macromolecular catalysts to drive an information ratchet. *Nat. Commun.* **10**, 3837 (2019).
42. Jayalath, I. M., Wang, H., Mantel, G., Kariyawasam, L. S. & Hartley, C. S. Chemically fueled transient geometry changes in diphenic acids. *Org. Lett.* **22**, 7567–7571 (2020).
43. Jayalath, I. M., Gerken, M. M., Mantel, G. & Hartley, C. S. Substituent effects on transient, carbodiimide-induced geometry changes in diphenic acids. *J. Org. Chem.* **86**, 12024–12033 (2021).
44. Amano, S. et al. Insights from an information thermodynamics analysis of a synthetic molecular motor. *Nat. Chem.* (2022) <https://doi.org/10.1038/s41557-022-00899-z>.
45. Ma, B. & Nussinov, R. Enzyme dynamics point to stepwise conformational selection in catalysis. *Curr. Opin. Chem. Biol.* **14**, 652–659 (2010).
46. Kamerlin, S. C. & Warshel, A. At the dawn of the 21<sup>st</sup> century: is dynamics the missing link for understanding enzyme catalysis? *Proteins* **78**, 1339–1375 (2010).
47. Krajnik, B. et al. Defocused imaging of UV-driven surface-bound molecular motors. *J. Am. Chem. Soc.* **139**, 7156–7159 (2017).
48. Roke, D., Wezenberg, S. J. & Feringa, B. L. Molecular rotary motors: unidirectional motion around double bonds. *Proc. Natl. Acad. Sci. USA* **115**, 9423–9431 (2018).

## Acknowledgements

We thank the Engineering and Physical Sciences Research Council (EPSRC; grant number EP/P027067/1), the European Research Council (ERC; Advanced Grant number 786630) and the German Research Foundation (DFG; Individual Postdoctoral Fellowship to E.K.) for funding, the University of Manchester’s Department of Chemistry Services for mass

spectrometry, I. J. Vitorica-Yrezabal for X-ray crystallography used to assign atropisomer handedness, S. Amano and J. M. Gallagher for useful discussions, and S. Jantzen of Biocinematics for the rotary motor animation. D.A.L. is a Royal Society Research Professor.

## Author information

### Author notes

1. These authors contributed equally: Stefan Borsley, Elisabeth Kreidt, Benjamin M. W. Roberts

### Affiliations

1. Department of Chemistry, University of Manchester, Manchester, UK

Stefan Borsley, Elisabeth Kreidt, David A. Leigh & Benjamin M. W. Roberts

2. School of Chemistry and Molecular Engineering, East China Normal University, Shanghai, China

David A. Leigh

### Contributions

S.B., E.K. and B.M.W.R. designed and carried out the experiments. D.A.L. directed the research. All authors contributed to the analysis of the results and the writing of the manuscript.

### Corresponding author

Correspondence to [David A. Leigh](#).

### Ethics declarations

## **Competing interests**

The authors declare no competing interests.

## **Peer review**

### **Peer review information**

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

## **Additional information**

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## **Supplementary information**

### **Supplementary Information**

Supplementary sections 1–7 and Appendix

### **Supplementary Video 1**

Animation of the design and autonomous chemically fuelled rotation of motor molecule **1a**. Credit: Stuart Jantzen, Biocinematics.

### **Peer Review File**

## **Rights and permissions**

### **Reprints and Permissions**

## **About this article**

## Cite this article

Borsley, S., Kreidt, E., Leigh, D.A. *et al.* Autonomous fuelled directional rotation about a covalent single bond. *Nature* **604**, 80–85 (2022).  
<https://doi.org/10.1038/s41586-022-04450-5>

- Received: 30 November 2021
- Accepted: 20 January 2022
- Published: 06 April 2022
- Issue Date: 07 April 2022
- DOI: <https://doi.org/10.1038/s41586-022-04450-5>

## Share this article

Anyone you share the following link with will be able to read this content:

[Get shareable link](#)

Sorry, a shareable link is not currently available for this article.

[Copy to clipboard](#)

Provided by the Springer Nature SharedIt content-sharing initiative

---

This article was downloaded by **calibre** from <https://www.nature.com/articles/s41586-022-04450-5>

- Article
- [Published: 06 April 2022](#)

# Multifunctional biocatalyst for conjugate reduction and reductive amination

- [Thomas W. Thorpe](#) [ORCID: orcid.org/0000-0001-6226-2936](#)<sup>1</sup>,
- [James R. Marshall](#)<sup>1</sup>,
- [Vanessa Harawa](#)<sup>1</sup>,
- [Rebecca E. Ruscoe](#)<sup>1</sup>,
- [Anibal Cuetos](#)<sup>2</sup>,
- [James D. Finnigan](#) [ORCID: orcid.org/0000-0002-8514-784X](#)<sup>3</sup>,
- [Antonio Angelastro](#) [ORCID: orcid.org/0000-0002-4023-7411](#)<sup>1</sup>,
- [Rachel S. Heath](#) [ORCID: orcid.org/0000-0003-1910-8268](#)<sup>1</sup>,
- [Fabio Parmeggiani](#) [ORCID: orcid.org/0000-0001-5861-9269](#)<sup>1</sup> nAff<sup>6</sup>,
- [Simon J. Charnock](#)<sup>3</sup>,
- [Roger M. Howard](#) [ORCID: orcid.org/0000-0001-5884-4896](#)<sup>4</sup>,
- [Rajesh Kumar](#)<sup>4</sup>,
- [David S. B. Daniels](#) [ORCID: orcid.org/0000-0002-9092-1377](#)<sup>5</sup>,
- [Gideon Grogan](#)<sup>2</sup> &
- [Nicholas J. Turner](#) [ORCID: orcid.org/0000-0002-8708-0781](#)<sup>1</sup>

[Nature](#) volume **604**, pages 86–91 (2022)

- 2433 Accesses
- 55 Altmetric
- [Metrics details](#)

## Subjects

- [Biocatalysis](#)
- [Synthetic chemistry methodology](#)
- [Oxidoreductases](#)

## Abstract

Chiral amine diastereomers are ubiquitous in pharmaceuticals and agrochemicals<sup>1</sup>, yet their preparation often relies on low-efficiency multi-step synthesis<sup>2</sup>. These valuable compounds must be manufactured asymmetrically, as their biochemical properties can differ based on the chirality of the molecule. Herein we characterize a multifunctional biocatalyst for amine synthesis, which operates using a mechanism that is, to our knowledge, previously unreported. This enzyme (EneIRED), identified within a metagenomic imine reductase (IRED) collection<sup>3</sup> and originating from an unclassified *Pseudomonas* species, possesses an unusual active site architecture that facilitates amine-activated conjugate alkene reduction followed by reductive amination. This enzyme can couple a broad selection of  $\alpha,\beta$ -unsaturated carbonyls with amines for the efficient preparation of chiral amine diastereomers bearing up to three stereocentres. Mechanistic and structural studies have been carried out to delineate the order of individual steps catalysed by EneIRED, which have led to a proposal for the overall catalytic cycle. This work shows that the IRED family can serve as a platform for facilitating the discovery of further enzymatic activities for application in synthetic biology and organic synthesis.

This is a preview of subscription content

## Access options

Subscribe to Nature+

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

\$29.99

monthly

[Subscribe](#)

Subscribe to Journal

Get full journal access for 1 year

\$199.00

only \$3.90 per issue

[Subscribe](#)

All prices are NET prices.

VAT will be added later in the checkout.

Tax calculation will be finalised during checkout.

Buy article

Get time limited or full article access on ReadCube.

\$32.00

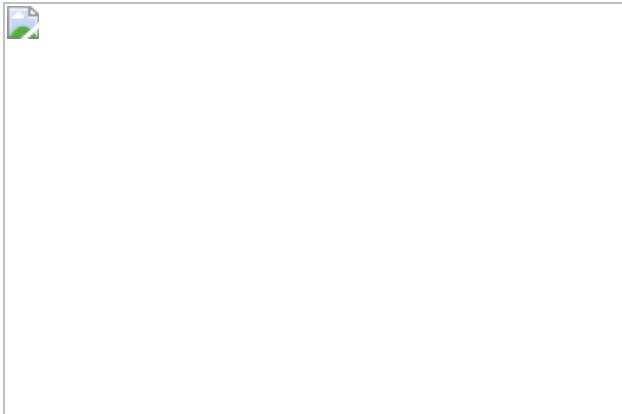
[Buy](#)

All prices are NET prices.

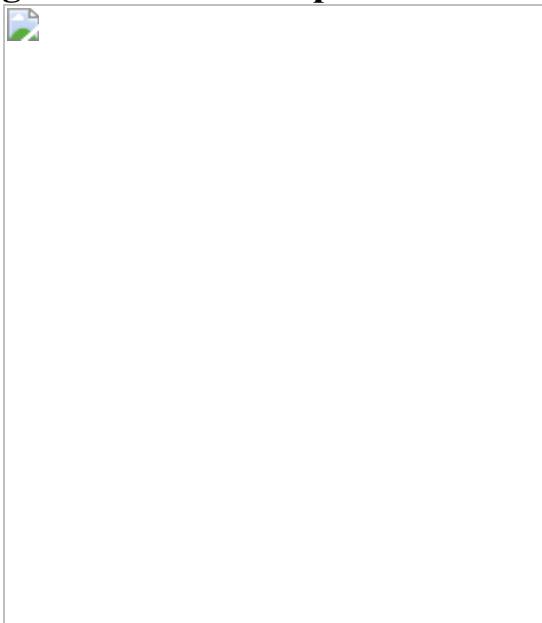
### **Additional access options:**

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

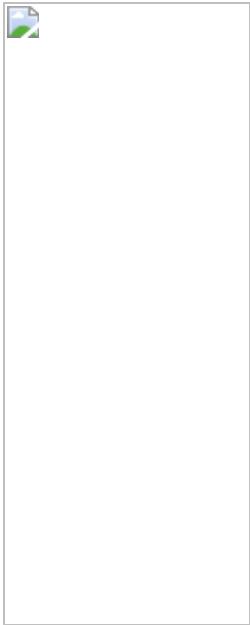
**Fig. 1: Enantioenriched amine diastereomers and one-pot enzymatic strategies for their synthesis.**



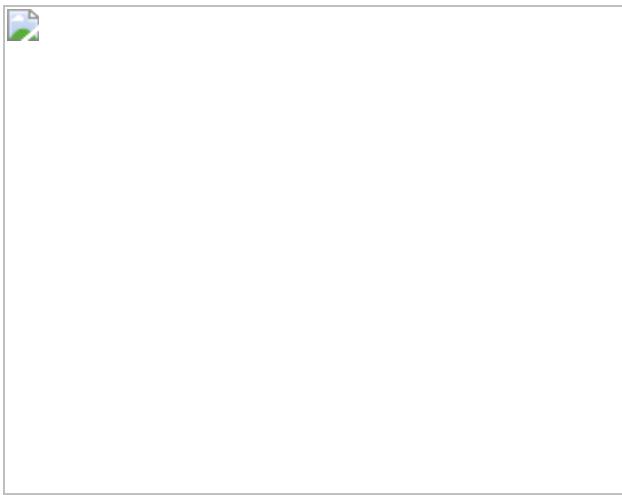
**Fig. 2: Substrate scope of EneIRED-catalysed CR-RA.**



**Fig. 3: Mechanistic and structural studies.**



**Fig. 4: Proposed catalytic cycle of productive EneIRED CR-RA and extension to six-electron CR-RA.**



## Data availability

The data supporting the findings of this study are available within the paper and its [Supplementary Information](#), and NMR traces are available from the Mendeley data repository (<https://data.mendeley.com>) at <https://doi.org/10.17632/fhc429t33c.1>. Sequence data have been deposited in Genbank (accession numbers MW854365, MW925135–MW925140) and the coordinate files and structure factors have been deposited in the PDB with accession number 7A3W.

## References

1. Jarvis, L. M. The new drugs of 2019. *Chem. Eng. News* **98**, 30–36 (2020).
2. Afanasyev, O. I., Kuchuk, E., Usanov, D. L. & Chusov, D. Reductive amination in the synthesis of pharmaceuticals. *Chem. Rev.* **119**, 11857–11911 (2019).
3. Marshall, J. R. et al. Screening and characterization of a diverse panel of metagenomic imine reductases for biocatalytic reductive amination. *Nat. Chem.* **13**, 140–148 (2021).
4. Roughley, S. D. & Jordan, A. M. The medicinal chemist’s toolbox: an analysis of reactions used in the pursuit of drug candidates. *J. Med. Chem.* **54**, 3451–3479 (2011).
5. Yasukawa, T., Masuda, R. & Kobayashi, S. Development of heterogeneous catalyst systems for the continuous synthesis of chiral amines via asymmetric hydrogenation. *Nat. Catal.* **2**, 1088–1092 (2019).
6. Wu, Z. et al. Secondary amines as coupling partners in direct catalytic asymmetric reductive amination. *Chem. Sci.* **10**, 4509–4514 (2019).
7. Skrypai, V., Varjosaari, S. E., Azam, F., Gilbert, T. M. & Adler, M. J. Chiral Brønsted acid-catalyzed metal-free asymmetric direct reductive amination using 1-hydrosilatrane. *J. Org. Chem.* **84**, 5021–5026 (2019).
8. Aleku, G. A. et al. A reductive aminase from *Aspergillus oryzae*. *Nat. Chem.* **9**, 961–969 (2017).
9. Mayol, O. et al. A family of native amine dehydrogenases for the asymmetric reductive amination of ketones. *Nat. Catal.* **2**, 324–333 (2019).

10. Yang, Y., Cho, I., Qi, X., Liu, P. & Arnold, F. H. An enzymatic platform for the asymmetric amination of primary, secondary and tertiary C( $sp^3$ )–H bonds. *Nat. Chem.* **11**, 987–993 (2019).
11. Li, T. et al. Efficient, chemoenzymatic process for, anufacture of the Boceprevir bicyclic [3.1.0]proline intermediate based on amine oxidase-catalyzed desymmetrization. *J. Am. Chem. Soc.* **134**, 6467–6472 (2012).
12. Zhou, J. & List, B. Organocatalytic asymmetric reaction cascade to substituted cyclohexylamines. *J. Am. Chem. Soc.* **129**, 7498–7499 (2007).
13. Monti, D. et al. Cascade coupling of ene-reductases and  $\omega$ -transaminases for the stereoselective synthesis of diastereomerically enriched amines. *ChemCatChem* **7**, 3106–3109 (2015).
14. France, S. P., Hepworth, L. J., Turner, N. J. & Flitsch, S. L. Constructing biocatalytic cascades: in vitro and in vivo approaches to de novo multi-enzyme pathways. *ACS Catal.* **7**, 710–724 (2017).
15. Huffman, M. A. et al. Design of an in vitro biocatalytic cascade for the manufacture of islatravir. *Science* **366**, 1255–1259 (2019).
16. Toogood, H. S. & Scrutton, N. S. Discovery, Characterization, engineering, and applications of ene-reductases for industrial biocatalysis. *ACS Catal.* **8**, 3532–3549 (2018).
17. Roth, S., Kilgore, M. B., Kutchan, T. M. & Müller, M. Exploiting the catalytic diversity of short-chain dehydrogenases/reductases: versatile enzymes from plants with extended imine substrate scope. *ChemBioChem* **19**, 1849–1852 (2018).
18. Hyslop, J. F. et al. Biocatalytic synthesis of chiral N-functionalized amino acids. *Angew. Chemie. Int. Ed.* **57**, 13821–13824 (2018).
19. Kato, Y., Yamada, H. & Asano, Y. Stereoselective synthesis of opine-type secondary amine carboxylic acids by a new enzyme opine

- dehydrogenase use of recombinant enzymes. *J. Mol. Catal. B* **1**, 151–160 (1996).
20. Schober, M. et al. Chiral synthesis of LSD1 inhibitor GSK2879552 enabled by directed evolution of an imine reductase. *Nat. Catal.* **2**, 909–915 (2019).
  21. Bornadel, A. et al. Technical considerations for scale-up of imine-reductase-catalyzed reductive amination: a case study. *Org. Process Res. Dev.* **23**, 1262–1268 (2019).
  22. Hussain, S. et al. An (*R*)-imine reductase biocatalyst for the asymmetric reduction of cyclic imines. *ChemCatChem* **7**, 579–583 (2015).
  23. Yao, P., Xu, Z., Yu, S., Wu, Q. & Zhu, D. Imine reductase-catalyzed enantioselective reduction of bulky  $\alpha,\beta$ -unsaturated imines en route to a pharmaceutically important morphinan skeleton. *Adv. Synth. Catal.* **361**, 556–561 (2019).
  24. Mitsukura, K. et al. Purification and characterization of a novel (*R*)-imine reductase from *Streptomyces* sp. GF3587. *Biosci. Biotechnol. Biochem.* **75**, 1778–1782 (2011).
  25. Lenz, M. et al. Asymmetric ketone reduction by imine reductases. *ChemBioChem* **18**, 253–256 (2017).
  26. Thorpe, T. W. et al. One-pot biocatalytic cascade reduction of cyclic enimines for the preparation of diastereomerically enriched *N*-heterocycles. *J. Am. Chem. Soc.* **141**, 19208–19213 (2019).
  27. Steiningerova, L. et al. Different reaction specificities of  $F_{420}H_2$ -dependent reductases facilitate pyrrolobenzodiazepines and lincomycin to fit their biological targets. *J. Am. Chem. Soc.* **142**, 3440–3448 (2020).
  28. Trenti, F. et al. Early and late steps of quinine biosynthesis. *Org. Lett.* **23**, 1793–1797 (2021).

29. Mitsukura, K., Suzuki, M., Tada, K., Yoshida, T. & Nagasawa, T. Asymmetric synthesis of chiral cyclic amine from cyclic imine by bacterial whole-cell catalyst of enantioselective imine reductase. *Org. Biomol. Chem.* **8**, 4533–4535 (2010).
30. France, S. P. et al. Identification of novel bacterial members of the imine reductase enzyme family that perform reductive amination. *ChemCatChem* **10**, 510–514 (2018).
31. Mangas-Sánchez, J. et al. Asymmetric synthesis of primary amines catalyzed by thermotolerant fungal reductive aminases. *Chem. Sci.* **11**, 5052–5057 (2020).
32. Montgomery, S. L. et al. Characterization of imine reductases in reductive amination for the exploration of structure–activity relationships. *Sci. Adv.* **6**, 9320 (2020).
33. Ouellet, S. G., Walji, A. M. & Macmillan, D. W. C. Enantioselective organocatalytic transfer hydrogenation reactions using Hantzsch esters. *Acc. Chem. Res.* **40**, 1327–1339 (2007).
34. Santi, N., Morrill, L. C., Świderek, K., Moliner, V. & Luk, L. Y. P. Transfer hydrogenations catalyzed by streptavidin-hosted secondary amine organocatalysts. *Chem. Commun.* **57**, 1919–1922 (2021).
35. Rodríguez-Mata, M. et al. Structure and activity of NADPH-dependent reductase Q1EQE0 from *Streptomyces kanamyceticus*, which catalyses the *R*-selective reduction of an imine substrate. *ChemBioChem* **14**, 1372–1379 (2013).
36. Holm, L. Benchmarking fold detection by DaliLite v.5. *Bioinformatics* **35**, 5326–5327 (2019).
37. Lenz, M. et al. New imine-reducing enzymes from β-hydroxyacid dehydrogenases by single amino acid substitutions. *Protein Eng. Des. Sel.* **31**, 109–120 (2018).

38. Huber, T. et al. Direct reductive amination of ketones: structure and activity of *S*-selective imine reductases from *Streptomyces*. *ChemCatChem* **6**, 2248–2252 (2014).
39. Man, H. et al. Structure, activity and stereoselectivity of NADPH-dependent oxidoreductases catalysing the *S*-selective reduction of the imine substrate 2-methylpyrroline. *ChemBioChem* **16**, 1052–1059 (2015).
40. Aleku, G. A. et al. Stereoselectivity and structural characterization of an imine reductase (IRED) from *Amycolatopsis orientalis*. *ACS Catal.* **6**, 3880–3889 (2016).
41. Trott, O. & Olson, A. J. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J. Comput. Chem.* **31**, 455–461 (2009).
42. Letunic, I. & Bork, P. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Res.* **49**, W293–W296 (2021).
43. Kabsch, W. XDS. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66**, 125–132 (2010).
44. Evans, P. Scaling and assessment of data quality. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **62**, 72–82 (2006).
45. Winter, G. xia2: an expert system for macromolecular crystallography data reduction. *J. Appl. Crystallogr.* **43**, 186–190 (2010).
46. Vagin, A. & Teplyakov, A. MOLREP: an automated program for molecular replacement. *J. Appl. Crystallogr.* **30**, 1022–1025 (1997).
47. Sharma, M. et al. A mechanism for reductive amination catalyzed by fungal reductive aminases. *ACS Catal.* **8**, 11534–11541 (2018).

48. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D* **60**, 2126–2132 (2004).
49. Murshudov, G. N., Vagin, A. A. & Dodson, E. J. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D* **53**, 240–255 (1997).

## Acknowledgements

T.W.T. is grateful for a CASE award from the UK Biotechnology and Biological Sciences Research Council (BBSRC) and Pfizer (BB/M011208/1). J.R.M. acknowledges a CASE award from the Industrial Biotechnology Innovation Centre (IBioIC), BBSRC and Prozomix Ltd. A.C. was funded by grant BB/P005578/1 from the BBSRC. N.J.T. is grateful to the ERC for the award of an Advanced Grant (742987). We thank J. P. Turkenburg and S. Hart for assistance with X-ray data collection and the Diamond Light Source for access to beamline I03 under proposal number mx-9948.

## Author information

### Author notes

1. Fabio Parmeggiani

Present address: Department of Chemistry, Materials and Chemical Engineering “G. Natta”, Politecnico di Milano, Milan, Italy

## Affiliations

1. Department of Chemistry, University of Manchester, Manchester Institute of Biotechnology, Manchester, UK

Thomas W. Thorpe, James R. Marshall, Vanessa Harawa, Rebecca E. Ruscoe, Antonio Angelastro, Rachel S. Heath, Fabio Parmeggiani & Nicholas J. Turner

2. Department of Chemistry, University of York, York, UK

Anibal Cuetos & Gideon Grogan

3. Prozomix, Haltwhistle, UK

James D. Finnigan & Simon J. Charnock

4. Pfizer Worldwide Research and Development, Groton, CT, USA

Roger M. Howard & Rajesh Kumar

5. Pfizer Worldwide Research and Development, Sandwich, UK

David S. B. Daniels

## Contributions

N.J.T., G.G., R.M.H., R.K. and D.S.B.D. devised and supervised the project. F.P. and R.E.R. managed the project. T.W.T. and A.A. performed mechanistic studies. T.W.T. and R.E.R. carried out substrate scope reactions. T.W.T. and V.H. carried out preparative scale reactions. T.W.T., R.E.R and V.H. synthesized substrates and standards. A.C., T.W.T. and G.G. performed crystallographic and docking studies. T.W.T. and R.S.H. undertook site-directed mutagenesis. J.R.M., S.J.C. and J.D.F. performed genetic identification, cloning and bioinformatics. T.W.T. and J.R.M. produced and purified the biocatalyst. N.J.T., G.G., R.M.H., R.K., D.S.B.D., S.J.C., F.P., J.D.F., A.C., R.E.R., V.H., J.R.M. and T.W.T. wrote the manuscript and generated the figures.

## Corresponding author

Correspondence to [Nicholas J. Turner](#).

## Ethics declarations

## Competing interests

The authors declare no competing interests.

## Peer review

### Peer review information

*Nature* thanks Dominic Campopiano, Sandy Schmidt and Thomas Ward for their contribution to the peer review of this work. Peer reviewer reports are available.

## Additional information

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## Extended data figures and tables

### [Extended Data Fig. 1 Phylogenetic IRED tree mapped against the reaction profiles of IRED-catalysed reduction of ene-imine I.](#)

Although the majority of the IREDs catalysed conventional imine reduction only, a small number were able to reduce both C=C and C=N bonds. Of these, EneIRED (pIR-120) possessed the highest propensity to forming the desired amine **II**.

### [Extended Data Fig. 2 Optimization of EneIRED-catalysed CR–RA reaction conditions.](#)

Conversion to the products of CR and CR–RA were elevated in glycine-OH pH 9.0, at moderate DMSO cosolvent concentration and at higher equivalents of amine donor. Formation of the direct RA product was not observed under any conditions.

### Extended Data Fig. 3 Scaled-up examples of EneIRED-catalysed CR-RA.

Several secondary and tertiary amines could be prepared including an example at elevated enone concentration and lower amine equivalents.

### Extended Data Fig. 4 Control reactions and isolated reactions of potential CR-RA pathway intermediates in EneIRED-catalysed CR-RA.

**a**, EneIRED CR-RA of **15** and **b** with NADPH. **b**, No enzyme control reaction. **c**, No recycling system control reaction. **d**, Reactions of potential CR-RA intermediate **15b** with NADP<sup>+</sup> or NADPH. **e**, Reaction of potential CR-RA intermediate **15'** with **b** using EneIRED. **f**, No amine control reactions with EneIRED point variants.

### Extended Data Fig. 5 Time-course studies of the CR-RA of **15** with **b** catalysed by wild-type EneIRED and point variants.

Both EneIRED-Y177A and EneIRED-Y181A exhibited a reduction in the rate of CR and CR-RA product formation compared to wild-type EneIRED, indicating that both residues are important for efficient catalysis. Notably, for EneIRED-Y177A the concentration of the ketone intermediate was comparatively low throughout the reaction, suggesting that Y177 is more important for CR than RA.

### Extended Data Fig. 6 Active site of EneIRED highlighting electron density.

**a**, Side chains, with density corresponding to the refined  $2F_o - F_c$  map (blue) at a level of  $1\sigma$ . **b**, NADP<sup>+</sup>, with density corresponding to the  $F_o - F_c$  difference map (green) at a level of  $3\sigma$  obtained from refinement in the absence of the ligand, with refined atoms included for clarity.  $F_o$  and  $F_c$  stand for the observed and calculated structure factor amplitudes, respectively.

## **Extended Data Table 1 Data collection and refinement statistics (molecular replacement) for EneIRED in complex with NADP<sup>+</sup>**

## **Supplementary information**

### **Supplementary Information**

This file contains Supplementary Text, Supplementary Figures, Supplementary Discussion and Supplementary References

### **Peer Review File**

## **Rights and permissions**

### **Reprints and Permissions**

## **About this article**

### **Cite this article**

Thorpe, T.W., Marshall, J.R., Harawa, V. *et al.* Multifunctional biocatalyst for conjugate reduction and reductive amination. *Nature* **604**, 86–91 (2022). <https://doi.org/10.1038/s41586-022-04458-x>

- Received: 19 June 2021
- Accepted: 20 January 2022
- Published: 06 April 2022
- Issue Date: 07 April 2022
- DOI: <https://doi.org/10.1038/s41586-022-04458-x>

### **Share this article**

Anyone you share the following link with will be able to read this content:

[Get shareable link](#)

Sorry, a shareable link is not currently available for this article.

[Copy to clipboard](#)

Provided by the Springer Nature SharedIt content-sharing initiative

---

This article was downloaded by **calibre** from <https://www.nature.com/articles/s41586-022-04458-x>

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

[Download PDF](#)

- Article
- [Published: 08 February 2022](#)

# Automated iterative $Csp^3$ -C bond formation

- [Daniel J. Blair](#) [ORCID: orcid.org/0000-0002-2279-7538<sup>1</sup>](#),
- [Sriyankari Chitti](#) [ORCID: orcid.org/0000-0001-9835-2773<sup>1</sup>](#),
- [Melanie Trobe<sup>1</sup>](#),
- [David M. Kostyra<sup>1</sup>](#),
- [Hannah M. S. Haley<sup>1</sup>](#),
- [Richard L. Hansen<sup>2</sup>](#),
- [Steve G. Ballmer<sup>2</sup>](#),
- [Toby J. Woods](#) [ORCID: orcid.org/0000-0002-1737-811X<sup>3</sup>](#),
- [Wesley Wang](#) [ORCID: orcid.org/0000-0002-1420-3031<sup>1</sup>](#),
- [Vikram Mubayi<sup>1</sup>](#),
- [Michael J. Schmidt<sup>1</sup>](#),
- [Robert W. Pipal<sup>1</sup>](#),
- [Greg F. Morehouse<sup>1</sup>](#),
- [Andrea M. E. Palazzolo Ray<sup>1</sup>](#),
- [Danielle L. Gray](#) [ORCID: orcid.org/0000-0003-0059-2096<sup>3</sup>](#),
- [Adrian L. Gill<sup>2</sup>](#) &
- [Martin D. Burke](#) [ORCID: orcid.org/0000-0001-7963-7140<sup>1,4,5,6</sup>](#)

*Nature* volume 604, pages 92–97 (2022)

- 12k Accesses
- 104 Altmetric
- [Metrics details](#)

## Subjects

- [Catalysis](#)

- [Synthetic chemistry methodology](#).

## Abstract

Fully automated synthetic chemistry would substantially change the field by providing broad on-demand access to small molecules. However, the reactions that can be run autonomously are still limited. Automating the stereospecific assembly of  $Csp^3-C$  bonds would expand access to many important types of functional organic molecules<sup>1</sup>. Previously, methyliminodiacetic acid (MIDA) boronates were used to orchestrate the formation of  $Csp^2-Csp^2$  bonds and were effective building blocks for automating the synthesis of many small molecules<sup>2</sup>, but they are incompatible with stereospecific  $Csp^3-Csp^2$  and  $Csp^3-Csp^3$  bond-forming reactions<sup>3,4,5,6,7,8,9,10</sup>. Here we report that hyperconjugative and steric tuning provide a new class of tetramethyl  $N$ -methyliminodiacetic acid (TIDA) boronates that are stable to these conditions. Charge density analysis<sup>11,12,13</sup> revealed that redistribution of electron density increases covalency of the N–B bond and thereby attenuates its hydrolysis. Complementary steric shielding of carbonyl  $\pi$ -faces decreases reactivity towards nucleophilic reagents. The unique features of the iminodiacetic acid cage<sup>2</sup>, which are essential for generalized automated synthesis, are retained by TIDA boronates. This enabled  $Csp^3$  boronate building blocks to be assembled using automated synthesis, including the preparation of natural products through automated stereospecific  $Csp^3-Csp^2$  and  $Csp^3-Csp^3$  bond formation. These findings will enable increasingly complex  $Csp^3$ -rich small molecules to be accessed via automated assembly.

## Main

Automated iterative assembly of chemical building blocks broadens access to innovation at the molecular scale<sup>1</sup>. Methods for reversibly attenuating the reactivity of the functional group handles used to link such blocks are critical for these platforms. For unsaturated ( $Csp^2$ -rich) organic small molecules, such lego-like assembly in automated and/or manual fashion has been achieved by many different research groups worldwide using  $N$ -methyliminodiacetic acid (MIDA) boronates<sup>2,14,15,16</sup>, which are compatible with anhydrous basic  $Csp^2$  cross-coupling conditions (Fig. 1a). An important advantage of MIDA relative to other ligands that attenuate boronic acid reactivity (for example, 1,8-diaminonaphthalene, anthranilamide, ethanolamine, fluoride)<sup>17</sup> is that MIDA boronates display a tuneable affinity for silica gel, which permits generalized automated purification<sup>2</sup>. This unique feature enabled assembly of 14 distinct classes of small organic molecules using one automated process<sup>2</sup>.

**Fig. 1: Lego-like chemical synthesis.**

---

 **figure 1**

**a**, Iterative chemical synthesis. Automated synthesis is achieved using bifunctional MIDA boronate building blocks. Controlled removal of MIDA enables iterative synthesis. **b**, Leading  $Csp^3$ -C bond-formation methods.  $Csp^3$ - $Csp^2$  cross-coupling of organoboranes is typically achieved under aqueous basic conditions. 1,2-Metallate rearrangements of boronic esters achieve  $Csp^3$ - $Csp^3$  bond formation by using Grignard and organolithium reagents. **c**, Sensitivity of MIDA boronates to  $Csp^3$ -C bond-forming conditions. Conditions permissive of  $Csp^3$ -C bond formation cleave MIDA boronates and are therefore incompatible with automated lego-like synthesis. *p*-tol, *para*-toluene; TIBO, 2,4,6-triisopropylbenzoate.

Important areas of chemical space remain inaccessible with this first-generation platform, particularly for molecules rich in non-planar and potentially stereogenic  $sp^3$ -hybridized carbon atoms ( $Csp^3$ ). This represents a substantial limitation, because  $Csp^3$ -rich molecules constitute some of the most notable natural products, medicines<sup>18</sup>, biological probes and functional materials<sup>19</sup>. An important goal is thus to expand automated modular synthesis to include  $Csp^3$ -rich small molecules.

Many recent breakthroughs in stereospecific formation of  $Csp^3$ -C bonds through Suzuki–Miyaura couplings<sup>6,7</sup> and 1,2-metallate rearrangements<sup>8,9,10</sup> stand to enable advances in this direction (Fig. 1b). However, most of these reactions require either aqueous basic conditions that hydrolyse MIDA boronates, or nucleophilic reagents that react with MIDA boronates (Fig. 1c). In both cases loss of the MIDA protecting group will lead to uncontrolled couplings and form complex mixtures and/or oligomeric products. We thus sought hyperstable boronates to enable lego-like small-molecule synthesis via iterative  $Csp^3$ -C bond formation.

## Identifying a hyperstable boronate

Mechanistic studies on MIDA boronate hydrolysis provided a foundation for developing hyperstabilized variants<sup>20</sup>. There are two mechanisms for MIDA boronate hydrolysis. The first involves frustrated Lewis pair-like activation of water by the dative N–B bond. The second mechanism involves ester hydrolysis-like cleavage of a MIDA carbonyl (C=O) group by hydroxide. Steric shielding would probably protect the carbonyl carbons from hydroxide and nucleophiles, yet steric effects are known to activate frustrated Lewis pair behaviour of N–B bonds<sup>21,22,23</sup>. Although hydrolysis studies of MIDA boronates<sup>20</sup> indicated electronic tuning of the N–B bond could be achieved via modifying the organic group attached to boron, we required a building block-independent solution. So, at the outset it was unclear whether steric or electronic effects could be leveraged to create a more stable ligand.

Using [<sup>18</sup>O]water to probe the hydrolysis of MIDA boronate **1a**<sup>20</sup>, we first established the N–B bond as the primary hydrolysis mechanism under aqueous basic  $Csp^3$ – $Csp^2$  Suzuki–Miyaura coupling conditions (THF/H<sub>2</sub>O, K<sub>2</sub>CO<sub>3</sub>, 60 °C) (Fig. [2a](#)). We thus required a MIDA derivative that could suppress the frustrated Lewis pair like reactivity of the N–B bond. To better understand this behaviour, we used <sup>1</sup>H NMR to study the stability of a range of substituted MIDA derivatives (**1b**, **3**–**9**) in deuterated solvent (THF-d<sub>8</sub>/D<sub>2</sub>O, K<sub>2</sub>CO<sub>3</sub>, 60 °C) (Fig. [2b](#)).

**Fig. 2: TIDA boronates are exceptionally stable towards hydrolytic and nucleophilic cleavage.**

 [figure 2](#)

**a**, MIDA boronate **1a** is hydrolysed via N–B bond-mediated water activation under standard stereospecific  $Csp^3$  coupling conditions. **b**, Modifications of the MIDA ligand yield varying degrees of stability for the corresponding iminodiacetic acid boronates under common aqueous basic stereospecific  $Csp^3$ – $Csp^2$  coupling conditions. TIDA boronate **9** is exceptionally stable. eq., equivalents. **c**, TIDA boronates resist hydrolysis during  $Csp^3$ – $Csp^2$  Suzuki–Miyaura coupling with **10**, whereas MIDA boronates are completely hydrolysed. **d**, TIDA boronates were stable to the Grignard reagent  $iPrMgCl\cdot LiCl$  enabling  $Csp^3$ – $Csp^3$  bond-forming 1,2-metallate rearrangements, whereas MIDA boronates were cleaved under these conditions. **e**, The remarkable stability of TIDA boronates extends to  $tBuLi$ . Diastereospecificity (d.s.) = (final diastereoisomeric ratio/initial diastereoisomeric ratio)  $\times 100$ . dan, 1,8-diaminonaphthalene; pin, pinacolato.

Consistent with earlier precedent<sup>22,23</sup>, bulky groups on nitrogen (Fig. [2b3–5](#) and Supplementary Fig. [2a](#)) increased the rate of hydrolysis relative to MIDA (**1b**), probably increasing N–B frustrated Lewis pair-like behaviour. Remarkably, appending two *n*-butyl groups (**6**) to the iminodiacetic acid backbone caused almost no change in hydrolysis rate relative to MIDA (**1b**). Reducing the size of these substituents to ethyl groups (**7**) and methyl groups (**8**) provided impressive stabilization. Finally, we prepared a boronate derived from a highly sterically hindered tetramethylated variant of *N*-methyliminodiacetic acid **9** (TIDA) and, surprisingly, found it was highly stable under aqueous basic  $Csp^3$ – $Csp^2$  cross-coupling conditions, with more than 99% remaining after 6 h (Fig. [2b](#)).

Additional hydrolysis studies of TIDA boronate **21** under these conditions in protic solvent (Supplementary Figs. [3](#), [4](#)), [ $^{18}O$ ] labelling (Supplementary Fig. [5](#)) and neutral hydrolysis (Supplementary Fig. [6](#)) indicated that TIDA boronates were cleaved via N–B bond water activation. The hyperstable TIDA boronate was retained during  $Csp^3$ – $Csp^2$  Suzuki–Miyaura reaction between  $Csp^3$  boronate **10** and bifunctional halo-TIDA boronate **11b** to provide **12**, whereas MIDA boronate **11a** (Fig. [2c](#)) and related *N*-2-benzyloxycyclopentyliminodiacetic acid (BIDA) boronate **SI-14** (Supplementary Fig. [2b](#)) gave no desired product.

Encouraged by these results, we tested the stability of TIDA boronates to  $iPrMgCl\cdot LiCl$ , which promotes stereospecific  $Csp^3$ – $Csp^3$  bond-forming 1,2-metallate rearrangements<sup>10</sup> (Fig. [2d](#)). TIDA boronate **14b** resists cleavage by  $iPrMgCl\cdot LiCl$  to form the target product **15** in high yield, whereas MIDA boronate **14a** is cleaved under these conditions. Remarkably, TIDA boronates even tolerated highly reactive  $tBuLi$ , enabling the formation of **18** and **19** with excellent diastereocontrol (Fig. [2e](#)). Providing an additional practical advantage, our bifunctional sulfoxide<sup>10</sup> TIDA boronate building blocks (that is, **14**, **16** and **17**) are easily handled, bench-stable solids (Supplementary Fig. [1](#)).

## X-ray crystallographic studies

The stability of **9** towards aqueous base is surprising considering the strong precedent for increased reactivity of frustrated Lewis pairs derived from tetramethylpiperidine<sup>21</sup>. Hints at the origin of this stability were found on X-ray crystallographic analysis of single crystals of MIDA boronate **1c** (Fig. **3a**, inset) and TIDA boronate **20** (Fig. **3b**, inset), which revealed a torsional shift of greater than 10° along the N–B axis in TIDA **20** relative to MIDA **1c**.

**Fig. 3: Steric and electronic effects collaborate to stabilize TIDA boronates.**

 figure 3



**a**, Examination of non-spherical (that is, bonding) electron density ( $\rho$ ) in the plane of the iminodiacetic acid ring of MIDA boronate **1c**. Contour level  $0.0675 \text{ e } \text{\AA}^{-3}$ . Shown inset is the perspective looking down from N to B. **b**, Non-spherical electron density in the plane of the iminodiacetic acid ring of TIDA boronate **12** shows substantial electronic redistribution compared to MIDA boronate **1c**, particularly for the N–B bond. An associated 12° torsional shift increases hyperconjugative interactions along the N–B axis compared to MIDA **1c**. **c**, Comprehensive steric shielding of all four  $\pi$ -faces of TIDA boronates suppresses nucleophilic attack on the carbonyl carbons. **d**, Despite their high stability towards  $Csp^3$  bond-forming reactions, TIDA boronates are easily removed under aqueous or protic conditions at elevated temperatures. Contour plots were generated using VESTA 3 (ref. [42](#)).

Torsional effects can substantially influence<sup>24</sup> the magnitude of hyperconjugative stabilization (that is, staggered versus eclipsed ethane<sup>25</sup>). Torsional shifts in **20** bring three donor N–C bonds nearly antiperiplanar to three acceptor bonds (two B–O bonds and one B–C bond), potentially elevating hyperconjugation across the N–B bond. Backbone methylation would probably increase N–C donor ability, and internal angle compression in **20** (approximately 5°) suggests thermodynamic Thorpe–Ingold effects make the framework more rigid (Extended Data Fig. [1a](#)). Both effects probably reinforce putative hyperconjugation across the N–B bond in **20**. We thus questioned whether stabilizing hyperconjugative interactions across the N–B bond in TIDA boronates drives a reduction in the rate of N–B hydrolysis.

## Electron distribution analysis

To probe electronic effects experimentally, we performed quantum theory of atoms in molecules (QTAIM)-based charge density analysis<sup>[11,12,13](#)</sup> on X-ray crystal structures of MIDA boronate **1c** and TIDA boronate **20**. Multiple lines of evidence revealed that the dramatic reduction in hydrolysis for TIDA boronates is attributable to hyperconjugation-mediated redistribution of electron density that increases the covalency of the N–B bond.

Topology maps of crystallographically determined bonding electron density,  $\rho$ , demonstrated electron redistribution across planar slices spanning the iminodiacetic acid rings of MIDA **1c** and TIDA **20** (Fig. [3a,b](#)). Notable features of TIDA **20** included increased electron density spanning the N–B interatomic space (Fig. [3b](#) and Extended Data Fig. [2j](#)), and formation of a contiguous ring of electron density around the iminodiacetic acid cage (Fig. [3b](#)). The stabilizing nature of electronic redistribution with the N–B bond of TIDA **20** was supported by a negative Laplacian of electron density  $\nabla^2\rho(r)$  at boron and nitrogen valence shell charge concentrations (Extended Data Fig. [3j](#)). TIDA **20** therefore possesses an N–B bond of substantially increased covalent character.

Elongation of nitrogen-attached donor bonds and boron-attached acceptor bonds (Extended Data Fig. [1b](#)), consistent with previous studies of anomeric<sup>26</sup> and gauche effects<sup>27</sup>, supports threefold hyperconjugation along the N–B linkage. Anticipated increased  $\pi$ -character manifested localized increased ellipticity ( $\varepsilon$ )<sup>[11](#)</sup> (Extended Data Fig. [4c,f,g,j,l,m](#)), and redistributed electron density was supported by changes in  $\nabla^2\rho(r)$  (Extended Data Fig. [3e,f,k](#))<sup>28</sup>. Electrostatic potential maps, reduced polarization of the N–B bond<sup>29</sup> and  $^{11}\text{B}/^{13}\text{C}$  NMR shifts were also consistent with electron redistribution (Extended Data Fig. [5](#)). Additional stabilizing electronic redistributions were found on examination of  $\rho$ ,  $\nabla^2\rho(r)$  and  $\varepsilon$  surrounding boron-attached oxygens O1 and O4, which revealed more equal distribution of electron

density directed towards boron and the carbonyl carbons for TIDA **20** compared to MIDA **1c** (Extended Data Figs. [2b](#), [c](#), [h](#), [i](#), [3b](#), [c](#), [h](#), [i](#), [4c](#), [d](#), [i](#), [j](#) and [6](#)).

Increased electron sharing<sup>30</sup> is consistent with reduced propensity for frustrated Lewis pair activity<sup>[31,32,33](#)</sup>, and rationalizes the increased robustness of TIDA boronates towards N–B bond hydrolysis. Remarkably, these stark differences in N–B bond character are contrasted with similar bond lengths (MIDA **1c**: 1.6613(7) Å; TIDA **20**: 1.6632(5) Å).

## Steric shielding of TIDA boronates

Crystallographic data for TIDA **20** also indicated that the stability of TIDA boronates towards carbon nucleophiles (*i*PrMgCl·LiCl and *t*BuLi) arises from shielding of all four  $\pi$ -faces of the carbonyls by the attached methyl groups (Fig. [3c](#) and Extended Data Fig. [7a](#)<sup>34</sup>). Comparison with organolithium-stable Beak-type benzoates<sup>35</sup> revealed that **20** mirrors Beak-like shielding interactions via a transannular methyl group spanning the back face of the iminodiacetic acid cage (Extended Data Fig. [7b](#)).

## Synthetic utility of TIDA boronates

Reversible ligation is a requirement for deploying TIDA boronates in iterative cross-coupling-based building block assembly. An orthogonal pathway for hydrolysis involving C=O attack (Supplementary Fig. [7](#) and Extended Data Fig. [8a](#)) enabled TIDA boronates (**21**) to be deprotected to boronic acids (**22**), trifluoroboronate salts (**23**), and boronic esters<sup>36</sup> (**24**, **25**) by simply using aqueous basic or protic conditions at elevated temperature (Fig. [3d](#)).

TIDA boronates retain all other key features of their MIDA boronate counterparts that enabled automated building block-based synthesis<sup>2</sup>: TIDA ligand is accessible on the kg scale (Extended Data Fig. [8b](#)), TIDA boronates are prepared from boronic acids under Dean–Stark conditions or by using a dehydrated form of TIDA (Extended Data Fig. [8c](#)), TIDA boronates are stable to a wide range of common cross-coupling reactions (that is, Stille, Suzuki, Sonogashira, Heck, photochemical; see Extended Data Fig. [9](#)) and chemical transformations (that is, oxidation, reduction, borylation, olefination; see Extended Data Fig. [10](#)). Representative TIDA boronates retain a tuneable affinity for silica gel, being minimally mobilized in Et<sub>2</sub>O and rapidly eluted in THF (Extended Data Fig. [8d](#)). This feature enables the TIDA boronate group to act as a tag for generalized and automated catch-and-release purification<sup>2</sup>. The heightened stability of TIDA boronates further enabled us to create a self-contained next-generation synthesis machine (Supplementary Figs. [8–11](#) and [13](#)) to enable automated iterative assembly of *Csp*<sup>3</sup> boronate building blocks (Fig. [4a,d](#)).

**Fig. 4: TIDA boronates enable automated assembly of  $Csp^3$  boronate building blocks.**

 figure 4

**a**, TIDA boronates enable iterative Suzuki–Miyaura cross-coupling. **b**, Assembly of  $Csp^3$  building blocks via  $Csp^3$ – $Csp^2$  Suzuki–Miyaura cross-coupling enabled by TIDA boronates. MIDA boronates universally provided no product. **c**, Automated

stereospecific  $Csp^3$  cross-coupling with TIDA boronate **32** enables lego-like synthesis of ieodomycin C. **d**, TIDA boronates enable iterative 1,2-metallate rearrangements. **e**, Assembly of  $Csp^3$  building blocks via  $Csp^3$ - $Csp^3$  bond-forming 1,2-metallate rearrangements enabled by TIDA boronates. **f**, Sequential automated stereospecific  $Csp^3$ - $Csp^3$  bond formation using TIDA boronates enables lego-like synthesis of sch725674. For detailed experimental procedures see Supplementary Information. Yields for automated synthesis shown in parentheses. d.r., diastereoisomeric ratio.

Having established an iterative synthesis platform (Fig. [4a](#)), we investigated a series of stereospecific  $Csp^3$ - $Csp^2$  cross-coupling reactions with bifunctional halo-TIDA boronates (Fig. [4b](#)). Across a range of carbonate-promoted aqueous/protic stereospecific  $Csp^3$  cross-coupling reactions<sup>6,7</sup>, MIDA boronates were fully hydrolysed, whereas the corresponding TIDA boronates universally provided the desired products in good yields both in manual and automated formats (**12**, **26**, **27** and **28**). The increased stability of TIDA boronates also permitted use of stronger bases KOH (**31**)<sup>5</sup> and Ag<sub>2</sub>O (**30**)<sup>4</sup>.

Leveraging automated  $Csp^3$ - $Csp^2$  couplings with TIDA boronates, we targeted a lego-like total synthesis of ieodomycin C<sup>37</sup> (Fig. [4c](#)). Building block **31** (97:3 enantiomeric ration (e.r.)) underwent automated stereospecific  $Csp^3$  cross-coupling with bifunctional TIDA boronate **32** to provide **33** in >95:5 e.r. and 50% isolated yield after automated purification. TIDA boronate enabled functional group interconversion followed by deprotection (**34**), and Suzuki–Miyaura cross-coupling with vinyl halide **35** furnished diene **36** and ieodomycin C after deprotection.

The tolerability of TIDA boronates to  $iPrMgCl\cdot LiCl$  enabled 1,2-metallate rearrangements to be executed with bifunctional sulfoxide-TIDA boronate building blocks in manual and automated formats (Fig. [4d](#)) to prepare a variety of  $Csp^3$ - $Csp^3$  bonds in excellent yields (Fig. [4e](#), **15** and **37–42**). A triply boron-selective reaction was also achieved in the diastereospecific preparation of **43** and **44**.  $Csp^2$  boronic esters were also effective (**45**, **46**).

Reactivity differences between unhindered/hindered boronic esters<sup>38,39</sup> suggested potential for iteration-enabling kinetic selectivity within  $Csp^3$ - $Csp^3$  bond formation. Accordingly, we investigated the lego-like automated synthesis of macrocyclic antifungal natural product sch725674 (ref. [40](#)) (Fig. [4f](#)). Demonstrating the advantage of our approach over previous strategies<sup>41</sup> to access sch725674, our bifunctional sulfoxide-TIDA boronate **14b** enabled recursive application of the same assembly chemistries to form  $Csp^3$ - $Csp^3$  bonds. Additionally, the inclusion of a TIDA boronate enabled this entire process of multiple building block assembly via iterative  $Csp^3$ - $Csp^3$  bond formation to be executed in a fully automated and uninterrupted fashion.

(Supplementary Figs. 14 and 15). *n*-Pentyl pinacol boronic ester **47** was subject to 1,2-metallate rearrangement with sulfoxide-boronate **14b** to afford the target TIDA boronate **48** in high stereocontrol (>95:5 e.r.). Automated deprotection to the corresponding pinacol boronic ester **49** was followed by automated boronic ester-selective reaction with sulfoxide **50** to provide the core carbon scaffold of sch725674 (**52**) after oxidation of bisboronate **51**. Deprotection (TBAF, **53**), oxidative modification, macro-lactonization and final deprotection furnished sch725674 in only seven steps from bench-stable building blocks.

## Data availability

X-ray crystal structure data are available free of charge on the Cambridge Crystallographic Data Centre under the following accession numbers: 4-bromophenylboronic acid MIDA ester **1c**: structure 2087874, multipole refinement 2087875; 4-bromophenylboronic acid TIDA ester **20**: structure 2087872, multipole refinement 2087873; 3-bromophenylboronic acid dimethyl-MIDA ester **SI-10**: structure 2087648; ethynylboronic acid TIDA ester **SI-47**: structure 2087715; *cis*-2-bromovinylboronic acid TIDA ester **SI-49**: structure 2087714; *trans*-2-bromovinylboronic acid TIDA ester **32**: structure 2087712; sulfinyl benzoate *anti*-**SI-25**: structure 2087716; TIDA anhydride: structure 2120500. All other data are available in the main text or supplementary materials.

## References

1. Trobe, M. & Burke, M. D. The molecular industrial revolution: automated synthesis of small molecules. *Angew. Chem. Int. Edn* **57**, 4192–4214 (2018).
2. Li, J. et al. Synthesis of many different types of organic small molecules using one automated process. *Science* **347**, 1221–1226 (2015).
3. Imao, D., Glasspoole, B. W., Laberge, V. S. & Crudden, C. M. Cross coupling reactions of chiral secondary organoboronic esters with retention of configuration. *J. Am. Chem. Soc.* **131**, 5024–5025 (2009).
4. Lehmann, J. W. et al. Axial shielding of Pd(II) complexes enables perfect stereoretention in Suzuki–Miyaura cross-coupling of Csp<sup>3</sup> boronic acids. *Nat. Commun.* **10**, 1263 (2019).
5. Mlynarski, S. N., Schuster, C. H. & Morken, J. P. Asymmetric synthesis from terminal alkenes by cascades of diboration and cross-coupling. *Nature* **505**, 386–390 (2013).

6. Ma, X., Murray, B. & Biscoe, M. R. Stereoselectivity in Pd-catalysed cross-coupling reactions of enantioenriched nucleophiles. *Nat. Rev. Chem.* **4**, 584–599 (2020).
7. Cherney, A. H., Kadunce, N. T. & Reisman, S. E. Enantioselective and enantiospecific transition-metal-catalyzed cross-coupling reactions of organometallic reagents to construct C–C bonds. *Chem. Rev.* **115**, 9587–9652 (2015).
8. Leonori, D. & Aggarwal, V. K. Lithiation-borylation methodology and its application in synthesis. *Acc. Chem. Res.* **47**, 3174–3183 (2014).
9. Sharma, H. A., Essman, J. Z. & Jacobsen, E. N. Enantioselective catalytic 1,2-boronate rearrangements. *Science* **374**, 752–757 (2021).
10. Casoni, G. et al.  $\alpha$ -Sulfinyl benzoates as precursors to Li and Mg carbenoids for the stereoselective iterative homologation of boronic esters. *J. Am. Chem. Soc.* **139**, 11877–11886 (2017).
11. Bader, R. F. W., Slee, T. S., Cremer, D. & Kraka, E. Description of conjugation and hyperconjugation in terms of electronic distributions. *J. Am. Chem. Soc.* **105**, 5061–5068 (1983).
12. Bader, R. F. W. Ed. *Atoms in Molecules—A Quantum Theory* (Oxford Univ. Press, 1990).
13. Koritsanszky, T. S. & Coppens, P. Chemical applications of X-ray charge-density analysis. *Chem. Rev.* **101**, 1583–1628 (2001).
14. Fujita, K., Matsui, R., Suzuki, T. & Kobayashi, S. Concise total synthesis of (−)-myxalamide A. *Angew. Chem. Int. Edn* **51**, 7271–7274 (2012).
15. Seo, K.-B., Lee, I.-H., Lee, J., Choi, I. & Choi, T.-L. A rational design of highly controlled Suzuki–Miyaura catalyst-transfer polycondensation for precision synthesis of polythiophenes and their block copolymers: marriage of palladacycle precatalysts with MIDA-boronates. *J. Am. Chem. Soc.* **140**, 4335–4343 (2018).
16. Angelone, D. et al. Convergence of multiple synthetic paradigms in a universally programmable chemical synthesis machine. *Nat. Chem.* **13**, 63–69 (2021).
17. Lennox, A. J. J. & Lloyd-Jones, G. C. Selection of boron reagents for Suzuki–Miyaura coupling. *Chem. Soc. Rev.* **43**, 412–443 (2014).

18. Lovering, F., Bikker, J. & Humblet, C. Escape from flatland: increasing saturation as an approach to improving clinical success. *J. Med. Chem.* **52**, 6752–6756 (2009).
19. Worch, J. C. et al. Stereochemical enrichment of polymer properties. *Nat. Rev. Chem.* **3**, 514–535 (2019).
20. Gonzalaez, J. A. et al. MIDA boronates are hydrolysed fast and slow by two different mechanisms. *Nat. Chem.* **8**, 1067–1075 (2016).
21. Stephan, D. W. & Erker, G. Frustrated Lewis pairs: metal-free hydrogen activation and more. *Angew. Chem. Int. Edn* **49**, 46–76 (2010).
22. Mancilla, T. & Contreras, R. New bicyclic organylboronic esters derived from iminodiacetic acids. *J. Organomet. Chem.* **307**, 1–6 (1986).
23. Mancilla, T., de los Ángeles Calixto Romo, M. & Delgado, L. A. Synthesis and characterization of ( $N\rightarrow B$ ) phenyl[ $N$ -alkyl- $N$ -(2-alkyl)aminodiacetate- $O,O',N$ ]boranes and phenyl[ $N$ -alkyl- $N$ -(2-alkyl)aminodiacetate- $O,O',N$ ]boranes. *Polyhedron* **26**, 1023–1028 (2007).
24. Wu, J. I.-C. & von Ragué Schleyer, P. Hyperconjugation in hydrocarbons: not just a “mild sort of conjugation”. *Pure Appl. Chem.* **85**, 921–940 (2013).
25. Popovska, V. & Goodman, L. Hyperconjugation not steric repulsion leads to the staggered structure of ethane. *Nature* **411**, 565–568 (2001).
26. Senderowitz, H., Golender, L. & Fuchs, B. New supramolecular host systems. 2. 1,3,5,7-Tetraoxadecalin, 1,2-dimethoxyethane and the gauche effect reappraised. Theory vs. experiment. *Tetrahedron* **32**, 9707–9728 (1994).
27. Hoffman, R. W., Hrovat, D. A. & Borden, W. T. Is hyperconjugation responsible for the ‘gauche effect’ in 1-fluoropropane and other 2-substituted-1-fluoroethanes? *J. Chem. Soc. Perkin Trans. 2* **12**, 1719–1726 (1999).
28. Scherer, W. et al. Valence-shell charge concentrations and electron delocalization in alkylolithium complexes: negative hyperconjugation and agnostic bonding. *Chem. Eur. J.* **8**, 2324–2334 (2002).
29. Hirschfeld, F. L. Bonded-atom fragments for describing molecular charge densities. *Theor. Chim. Acta* **44**, 129–138 (1977).
30. Jonas, V., Frenking, G. & Reetz, M. T. Comparative theoretical study of Lewis acid–base complexes of  $BH_3$ ,  $BF_4$ ,  $BCl_3$ ,  $AlCl_3$ , and  $SO_2$ . *J. Am. Chem. Soc.* **116**,

8741–8753 (1994).

31. Skara, G., de Vleeschouwer, F., Geerlings, P., de Proft, F. & Pinter, B. Heterolytic splitting of molecular hydrogen by frustrated and classical Lewis pairs: a unified reactivity concept. *Sci. Rep.* **7**, 16024 (2017).
32. Schürmann, C. J. et al. Experimental charge density study on FLPs and a FLP reaction product. *Z. Kristallogr. Crystall. Mater.* **233**, 723–731 (2018).
33. Ullrich, M., Lough, A. J. & Stephan, D. W. Dihydrogen activation by B(p-C<sub>6</sub>F<sub>4</sub>H)<sub>3</sub> and phosphines. *Organometallics* **29**, 3647–3654 (2010).
34. Falivene, L. et al. Towards the online computer-aided design of catalytic pockets. *Nat. Chem.* **11**, 872–879 (2019).
35. Beak, P. & Carter, L. G. Dipole-stabilized carbanions from esters: α-oxo lithiations of 2,6-substituted benzoates of primary alcohols. *J. Org. Chem.* **46**, 2363–2373 (1981).
36. Landry, M. L., Hu, D. X., McKenna, G. M. & Burns, N. Z. Catalytic enantioselective dihalogenation and the selective synthesis of (−)-deschloromytilipin A and (−)-danicalipin A. *J. Am. Chem. Soc.* **138**, 5150–5158 (2016).
37. Mojid Mondol, M. A. et al. Ieodomycins A–D, antimicrobial fatty acids from a marine *Bacillus* sp. *J. Nat. Prod.* **74**, 1606–1612 (2011).
38. Blakemore, P. R., Marsden, S. P. & Vater, H. D. Reagent-controlled asymmetric homologation of boronic esters by enantioenriched main-group chiral carbenoids. *Org. Lett.* **8**, 773–776 (2006).
39. Roesner, S., Blair, D. J. & Aggarwal, V. K. Enantioselective installation of adjacent tertiary benzylic stereocentres using lithiation-borylation-protodeboronation methodology. Application to the synthesis of bifluranol and fluorohexestrol. *Chem. Sci.* **6**, 3718–3723 (2015).
40. Yang, S.-W. et al. Structure elucidation of sch725674 from *Aspergillus* sp. *J. Antibiot.* **58**, 535–538 (2005).
41. Fawcett, A. et al. Regio- and stereoselective homologation of 1,2-bis(boronic esters): stereocontrolled synthesis of 1,3-diols and sch725674. *Angew. Chem. Int. Edn* **55**, 14663–14667 (2016).

42. Momma, K. & Izumi, F. VESTA 3 for three-dimensional visualization of crystal, volumetric and morphology data. *J. Appl. Crystallogr.* **44**, 1272–1276 (2011).
43. O'Brien, N. J. et al. Synthesis, structure and reactivities of pentacoordinated phosphorus-boron bonded compounds. *Eur. J. Inorg. Chem.* **20**, 1995–2003 (2020).
44. Uno, B. E., Gillis, E. P. & Burke, M. D. Vinyl MIDA boronate: a readily accessible and highly versatile building block for small molecule synthesis. *Tetrahedron* **65**, 3130–3138 (2009).
45. Ma, Y. et al. Radical C–N borylation of aromatic amines enabled by a pyrylium reagent. *Chem. Eur. J.* **26**, 3738–3743 (2020).
46. Neuvonen, H., Neuvonen, K., Koch, A., Kleinpeter, E. & Pasanen, P. Electron-withdrawing substituents decrease the electrophilicity of the carbonyl carbon. An investigation with the aid of  $^{13}\text{C}$  NMR chemical shifts,  $\nu(\text{C}=\text{O})$  frequency values, charge densities, and isodesmic reactions to interpret substituent effects on reactivity. *J. Org. Chem.* **67**, 6995–7003 (2002).
47. Aspin, S., Goutierre, A.-S., Larini, P., Jazzaar, R. & Baudoin, O. Synthesis of aromatic  $\alpha$ -aminoesters: palladium-catalyzed long-range arylation of primary  $\text{Csp}^3\text{-H}$  bonds. *Angew. Chem. Int. Edn* **51**, 10808–10811 (2012).
48. Li, G., Ji, C.-L., Hong, X. & Szostak, M. Highly chemoselective, transition-metal-free transamidation of unactivated amides and direct amidation of alkyl esters by N-C/O-C cleavage. *J. Am. Chem. Soc.* **141**, 11161–11172 (2019).
49. Xie, X. & Stahl, S. S. Efficient and selective Cu/nitroxyl-catalyzed methods for aerobic oxidative lactonization of diols. *J. Am. Chem. Soc.* **137**, 3767–3770 (2015).
50. Yamamoto, Y., Nemoto, H., Kikuchi, R., Komatsu, H. & Suzuki, I. A conformationally rigid acyclic molecule. *J. Am. Chem. Soc.* **112**, 8598–8599 (1990).
51. Ueki, Y., Ito, H., Usui, I. & Breit, B. Formation of quaternary carbon centers by highly regioselective hydroformylation with catalytic amounts of a reversibly bound directing group. *Chem. Eur. J.* **17**, 8555–8558 (2011).

## Acknowledgements

We thank F. Sun and H. Yao at the UIUC Mass Spectrometry Facility, D. Olson, L. Zhu and N. Duay at the School of Chemical Sciences NMR Laboratory at UIUC for NMR services, and members of the Burke laboratory for discussions related to this project. The Bruker 500 MHz NMR spectrometer was obtained with the financial support of the Roy J. Carver Charitable Trust, Muscatine, Iowa, USA. D.J.B. thanks the Damon-Runyon Cancer Research Foundation for additional support during the COVID-19 pandemic. S. Denmark is acknowledged for manuscript review. We thank P.-J. Chen, Y. Tong, A. Blake, H. Auby and D. Szczepankiewicz for technical assistance. Support was provided by the following sources: M.D.B.: NIH (GM118185), NSF (CHE-1955838); D.J.B.: Illini 4000 Fellow of the Damon-Runyon Cancer Research Foundation DRG-2290-17; S.C.: ACS Division of Organic Chemistry Summer Undergraduate Research Fellowship, Henry Luce Foundation and the Illinois Scholars Undergraduate Research Program; M.B.T.: Erwin Schrödinger Post-Doctoral Fellow, Austrian Science Fund (FWF) (J3960-N34).

## Author information

### Affiliations

1. Roger Adams Laboratory, School of Chemical Sciences, University of Illinois at Urbana-Champaign, Urbana, IL, USA

Daniel J. Blair, Sriyankari Chitti, Melanie Trobe, David M. Kostyra, Hannah M. S. Haley, Wesley Wang, Vikram Mubayi, Michael J. Schmidt, Robert W. Pipal, Greg. F. Morehouse, Andrea M. E. Palazzolo Ray & Martin D. Burke

2. Department of Chemistry, REVOLUTION Medicines, Inc., Redwood City, CA, USA

Richard L. Hansen, Steve G. Ballmer & Adrian L. Gill

3. George L. Clark X-Ray Facility and 3M Materials Laboratory, University of Illinois at Urbana-Champaign, Urbana, IL, USA

Toby J. Woods & Danielle L. Gray

4. Carle Illinois College of Medicine, Urbana, IL, USA

Martin D. Burke

5. Arnold and Mabel Beckman Institute, University of Illinois at Urbana-Champaign, Urbana, IL, USA

Martin D. Burke

6. Department of Biochemistry, University of Illinois at Urbana-Champaign,  
Urbana, IL, USA

Martin D. Burke

## Contributions

The project was designed by D.J.B., M.J.S. and M.D.B. Experimental work was conducted by D.J.B., S.C., M.T., D.M.K., H.M.S.H., R.L.H., S.G.B., T.J.W., W.W., V.M., M.J.S., R.W.P., G.F.M., A.M.E.P.R. and D.L.G. Automated synthesis experiments were performed by D.J.B., M.T. and H.M.S.H with guidance from R.L.H. and S.G.B. Crystallographic data were collected by T.J.W. and D.L.G. Multipole structural refinements were performed by T.J.W. Small-molecule synthesis machines were designed, constructed and finalized by R.L.H., S.G.B., A.L.G. and M.D.B. D.J.B. and M.D.B. wrote the manuscript.

## Corresponding authors

Correspondence to [Daniel J. Blair](#) or [Martin D. Burke](#).

## Ethics declarations

## Competing interests

The University of Illinois has filed patent applications related to MIDA and TIDA boronates. M.D.B. is a founder, shareholder and consultant for REVOLUTION Medicines.

## Peer review

### Peer review information

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

## Additional information

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## Extended data figures and tables

### Extended Data Fig. 1 Structural consequences of backbone substitution on iminodiacetic acid boronates.

**a**, Internal angle compression from MIDA **1c** to dimethyl-MIDA **SI-10** to TIDA **20** suggest thermodynamic Thorpe–Ingold effects might further support hyperconjugative interactions by rigidifying the iminodiacetic acid cage of TIDA. **b**, Increases in nitrogen attached donor and boron attached acceptor bond lengths for TIDA **20** compared to MIDA **1c** support hyperconjugative transfer of electron density along the N-B bond (lengths in Å).

### Extended Data Fig. 2 Bond path total electron density for MIDA boronate **1c** and TIDA boronate **20**.

**a**, Legend and numbered molecular structures. **b–i** Follow anticlockwise around the iminodiacetic acid cage. **j–l**, The N-B, N-Me, and C-B bonds are also included. The maximum displayed value of the electron density (y-axis) is capped at  $4 \text{ eA}^{-3}$  to best capture the interatomic bonding region.

### Extended Data Fig. 3 Bond path Laplacian profiles comparing MIDA boronate **1c** and TIDA boronate **20**.

**a**, Legend and labelled structures. **b–i** Follow an anticlockwise path around MIDA **1c** and TIDA **20** as well as the N-B, N-Me, and C-B bonds (**j–l**).

### Extended Data Fig. 4 Bond path ellipticity profiles for MIDA boronate **1c** and TIDA boronate **20**.

**a**, Schematic representation of ellipticity ( $\epsilon$ ), which reflects an increase in directionality of electron density and is characteristic of increased  $\pi$ -character. **b**, Legend for sections **c–m**. Pronounced changes in ellipticity can be seen for TIDA boronate **20** compared to MIDA boronate **1c** particularly for the N-C2/3 (**f, g**), B-O1/4 (**c, j**), and N-B bonds, which are consistent with observations of redistributions of electron density around the N-B bond (see Fig. [3a, b](#)). Ellipticity profiles (**c–j**) follow an anticlockwise path around the iminodiacetic acid rings of **1c** and **20** starting with B-

O1 (**c**) and ending with O4-B (**j**). Profiles for the N-B, N-Me, and C-B bonds (**k-m**) are also included.

**Extended Data Fig. 5 Charge analysis of MIDA boronate 1c and TIDA boronate 20.**

**a**, Electrostatic potential surface for MIDA **1c**, showing side on and bottom perspectives. **b**, Electrostatic potential surface for TIDA **20**, showing side on and bottom perspectives. **c**, Stockholder charges comparing MIDA **1c** (black) and TIDA **20** (green), values shown are in electrons. **d**, Integrated charge comparing MIDA **1c** (black) and TIDA **20** (green), values shown are in electrons. TIDA backbone methyl groups were omitted for clarity on the charge plots. **e**, A downfield shift for TIDA boronates carbonyl carbons indicates a net electron depletion relative to MIDA boronates. Similarly, there is an up-field shift for the boron atoms in TIDA boronates relative to MIDA boronates indicating increased boronate complex-like character and elevated electronic shielding<sup>43,44,45</sup>. **f**, Trends in <sup>13</sup>C NMR carbonyl chemical shifts in CDCl<sub>3</sub> on sequential substitution of the  $\alpha$ -carbon with methyl groups and/or lactone formation and/or addition of an  $\alpha$ -dimethylamino group are provided for comparison<sup>46,47,48,49,50,51</sup>.

**Extended Data Fig. 6 Laplacian and electron density isosurfaces support redistributed density around the O4-B-O1 linkage of TIDA 20.**

**a**, The Laplacian at isosurface value -80 eA<sup>-5</sup> (shown in yellow) for MIDA **1c** indicates an isolated valence shell charge concentration (VSCC) at O4 (i.e. minimal lone-pair interactions). **b**, Unlike at O4 the Laplacian at O1 for MIDA **1c** reveals coalescence of lone-pair VSCC and the O1-C1 VSCC, pointing toward interaction between O1 and C1 (the adjacent carbonyl). **c**, End-on view of deformation density (isosurface value of 0.0034 eA<sup>-3</sup> in blue) down O4 in MIDA **1c** provides further evidence for charge localization at O4. **d**, In contrast to O4 the end-on view of deformation density down O1 for MIDA **1c** reveals the electron distribution along B-O1-C1, favours O1-C1. **e**, The Laplacian of TIDA **20** reveals interaction between lone-pair VSCC and both C4-O4/O4-B VSCCs. **f**, In contrast to MIDA **1c** the lone-pair VSCC of TIDA **20** at O1 coalesces with the B-O1 VSCC and not the O1-C1 VSCC. **g**, **h**, Consistent with these changes in VSCCs, deformation density at O4 (**g**) and O1 (**h**) for TIDA **20** supports electronic redistribution about the C4-O4-B-O1-C1 network compared to MIDA **1c**. Prepared using VESTA 3 (ref. <sup>42</sup>).

**Extended Data Fig. 7 Robust steric shielding suppresses carbonyl attack on TIDA boronates.**

**a**, Topological steric maps<sup>34</sup> of the plane perpendicular to the carbonyl carbons enable comparison of MIDA (**1c**), dimethyl-MIDA (**SI-10**), and TIDA (**20**). MIDA boronates (left column) experience minimal steric shielding, and methyl groups introduced in dimethyl-MIDA (A' and B') occupy pseudo-equatorial positions, minimally impacting carbonyl approach (centre column). The two additional methyl groups in TIDA (A and B) occupy pseudo-axial positions and establish transannular steric shielding interactions between A and the carbonyl on the opposite side of the TIDA framework (right column). Additionally, A and B shield their adjacent carbonyls towards approach at the Burgi-Dunitz angle. **b**, The transannular influence of methyl group A mirrors that of Beak-type 2,4,6-triisopropyl benzoates which are resistant to carbon centred nucleophiles.

**Extended Data Fig. 8 TIDA boronates retain all required properties to enable generalized automated synthesis.**

**a**, TIDA boronate **21** is hydrolysed by NaOH primarily *via* the ester hydrolysis mechanism. **b**, TIDA ligand and TIDA boronates can be prepared on scale. **c**, TIDA anhydride provides an alternative method to prepare TIDA boronates. **d**, TIDA boronates possess a binary affinity for silica gel, agnostic of the attached carbon fragment. They are minimally mobilized in Et<sub>2</sub>O and rapidly eluted in THF, enabling generalized and automatable catch-and-release purification. tol, *para*-toluene.

**Extended Data Fig. 9 TIDA boronates tolerate a diverse range of cross-coupling chemistry.**

**a**, Suzuki–Miyaura cross-coupling. **b**, Heck coupling. **c**, Sonogashira coupling. **d**, Photochemical Suzuki–Miyaura cross-coupling. **e**, Photochemical thioetherification. **f**, Stille coupling leading to bis-borylated dienes. **g**, Stille coupling leading to germylated dienes.

**Extended Data Fig. 10 Functional group interconversion of TIDA boronate building blocks.**

**a**, Ethynylboronic acid TIDA ester is readily converted into *E*- and *Z*- 2-bromovinylboronic acid TIDA esters with excellent stereocontrol. Images of X-ray crystal structures shown inset. **b**, Reduction of ethynylboronic acid TIDA ester furnishes vinylboronic acid TIDA ester which participates in epoxidation and Grubbs metathesis. **c**, Common functional group interconversion reactions are well tolerated by TIDA including oxidation, reduction, halogenation, reductive amination, Evans aldol and Takai olefination. **d–g**, A wide range of borylation chemistries are tolerated by TIDA boronates to produce mono-protected polyborylated building blocks,

including Miyaura (**d**), C–H borylation (**e**), diboration (**f**), and copper catalysed borylation (**g**).

## Supplementary information

### Supplementary Information

Procedures for chemical synthesis, automated synthesis and mechanistic data.

## Rights and permissions

### Reprints and Permissions

## About this article

### Cite this article

Blair, D.J., Chitti, S., Trobe, M. *et al.* Automated iterative  $Csp^3$ –C bond formation. *Nature* **604**, 92–97 (2022). <https://doi.org/10.1038/s41586-022-04491-w>

- Received: 20 September 2021
- Accepted: 28 January 2022
- Published: 08 February 2022
- Issue Date: 07 April 2022
- DOI: <https://doi.org/10.1038/s41586-022-04491-w>

### Share this article

Anyone you share the following link with will be able to read this content:

[Get shareable link](#)

Sorry, a shareable link is not currently available for this article.

[Copy to clipboard](#)

Provided by the Springer Nature SharedIt content-sharing initiative

This article was downloaded by **calibre** from <https://www.nature.com/articles/s41586-022-04491-w>

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

- Article
- [Published: 30 March 2022](#)

# A stable hippocampal code in freely flying bats

- [William A. Liberti III<sup>1</sup>](#) na1,
- [Tobias A. Schmid<sup>2</sup>](#) na1,
- [Angelo Forli<sup>1</sup>](#),
- [Madeleine Snyder<sup>1</sup>](#) &
- [Michael M. Yartsev](#) ORCID: orcid.org/0000-0003-0952-2801<sup>1,2</sup>

*Nature* volume **604**, pages 98–103 (2022)

- 6731 Accesses
- 181 Altmetric
- [Metrics details](#)

## Subjects

- [Hippocampus](#)
- [Neuroscience](#)

## Abstract

Neural activity in the hippocampus is known to reflect how animals move through an environment<sup>1,2</sup>. Although navigational behaviour may show considerable stability<sup>3,4,5,6</sup>, the tuning stability of individual hippocampal neurons remains unclear<sup>7,8,9,10,11,12</sup>. Here we used wireless calcium imaging to longitudinally monitor the activity of dorsal CA1 hippocampal neurons

in freely flying bats performing highly reproducible flights in a familiar environment. We find that both the participation and the spatial selectivity of most neurons remain stable over days and weeks. We also find that apparent changes in tuning can be largely attributed to variations in the flight behaviour of the bats. Finally, we show that bats navigating in the same environment under different room lighting conditions (lights on versus lights off) exhibit substantial changes in flight behaviour that can give the illusion of neuronal instability. However, when similar flight paths are compared across conditions, the stability of the hippocampal code persists. Taken together, we show that the underlying hippocampal code is highly stable over days and across contexts if behaviour is taken into account.

This is a preview of subscription content

## Access options

Subscribe to Nature+

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

\$29.99

monthly

[Subscribe](#)

Subscribe to Journal

Get full journal access for 1 year

\$199.00

only \$3.90 per issue

[Subscribe](#)

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

Buy article

Get time limited or full article access on ReadCube.

\$32.00

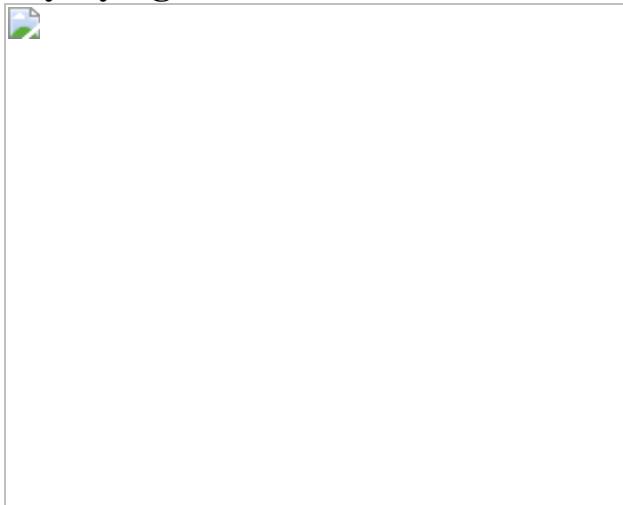
[Buy](#)

All prices are NET prices.

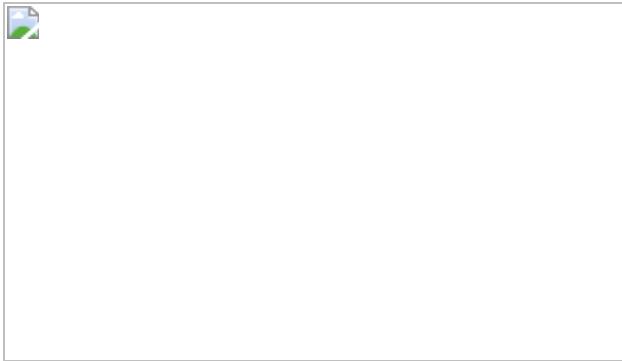
### **Additional access options:**

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

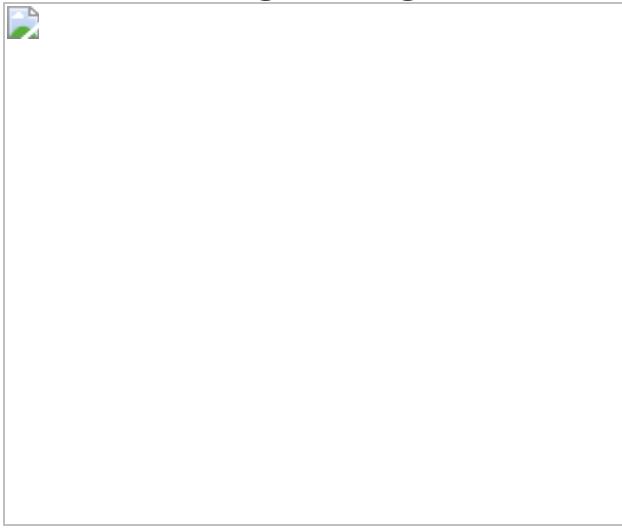
**Fig. 1: Wireless calcium imaging and spatial movement patterns in freely flying bats.**



**Fig. 2: Neural tuning for stable flight paths persists across days.**



**Fig. 3: Apparent changes in neural tuning can be explained by measurable changes in flight behaviour.**



**Fig. 4: Neural tuning persists despite a change in sensory context.**



## Data availability

The dataset from this study is available from the corresponding author on reasonable request. [Source data](#) are provided with this paper.

## Code availability

All analyses were conducted using custom code in MATLAB (MathWorks), which can be found at GitHub (<https://github.com/WALIII/ImBat>; <https://github.com/gardner-lab/FinchScope>).

## References

1. O'Keefe, J. & Dostrovsky, J. The hippocampus as a spatial map. Preliminary evidence from unit activity in the freely-moving rat. *Brain Res.* **34**, 171–175 (1971).
2. Moser, E. I., Moser, M.-B. & McNaughton, B. L. Spatial representation in the hippocampal formation: a history. *Nat. Neurosci.* **20**, 1448–1464 (2017).
3. Geva-Sagiv, M., Las, L., Yovel, Y. & Ulanovsky, N. Spatial cognition in bats and rats: from sensory acquisition to multiscale maps and navigation. *Nat. Rev. Neurosci.* **16**, 94–108 (2015).
4. Etienne, A. S. et al. Navigation through vector addition. *Nature* **396**, 161–164 (1998).
5. Harten, L., Katz, A., Goldshtein, A., Handel, M. & Yovel, Y. The ontogeny of a mammalian cognitive map in the real world. *Science* **369**, 194–197 (2020).
6. Toledo, S. et al. Cognitive map-based navigation in wild bats revealed by a new high-throughput tracking system. *Science* **369**, 188–193 (2020).
7. Ziv, Y. et al. Long-term dynamics of CA1 hippocampal place codes. *Nat. Neurosci.* **16**, 264–266 (2013).
8. Rubin, A., Geva, N., Sheintuch, L. & Ziv, Y. Hippocampal ensemble dynamics timestamp events in long-term memory. *eLife* **4**, e12247 (2015).

9. Gonzalez, W. G., Zhang, H., Harutyunyan, A. & Lois, C. Persistence of neuronal representations through time and damage in the hippocampus. *Science* **365**, 821–825 (2019).
10. Mankin, E. A., Diehl, G. W., Sparks, F. T., Leutgeb, S. & Leutgeb, J. K. Hippocampal CA2 activity patterns change over time to a larger extent than between spatial contexts. *Neuron* **85**, 190–201 (2015).
11. Thompson, L. T. & Best, P. J. Long-term stability of the place-field activity of single units recorded from the dorsal hippocampus of freely behaving rats. *Brain Res.* **509**, 299–308 (1990).
12. Hainmueller, T. & Bartos, M. Parallel emergence of stable and dynamic memory engrams in the hippocampus. *Nature* **558**, 292–296 (2018).
13. Moser, E. I., Kropff, E. & Moser, M.-B. Place cells, grid cells, and the brain’s spatial representation system. *Annu. Rev. Neurosci.* **31**, 69–89 (2008).
14. Payne, H. L., Lynch, G. F. & Aronov, D. Neural representations of space in the hippocampus of a food-caching bird. *Science* **373**, 343–348 (2021).
15. O’Keefe, J. & Nadel, L. *The Hippocampus as a Cognitive Map* (Clarendon Press, Oxford Univ. Press, 1978).
16. Yartsev, M. M. & Ulanovsky, N. Representation of three-dimensional space in the hippocampus of flying bats. *Science* **340**, 367–372 (2013).
17. Wohlgemuth, M. J., Yu, C. & Moss, C. F. 3D hippocampal place field dynamics in free-flying echolocating bats. *Front. Cell. Neurosci.* **12**, 270 (2018).
18. Tolman, E. C. Cognitive maps in rats and men. *Psychol. Rev.* **55**, 189–208 (1948).

19. Tsoar, A. et al. Large-scale navigational map in a mammal. *Proc. Natl Acad. Sci. USA* **108**, E718–E724 (2011).
20. Muller, R., Kubie, J. & Ranck, J. Spatial firing patterns of hippocampal complex-spike cells in a fixed environment. *J. Neurosci.* **7**, 1935–1950 (1987).
21. Lütcke, H., Margolis, D. J. & Helmchen, F. Steady or changing? Long-term monitoring of neuronal population activity. *Trends Neurosci.* **36**, 375–384 (2013).
22. Clopath, C., Bonhoeffer, T., Hübener, M. & Rose, T. Variance and invariance of neuronal long-term representations. *Philos. Trans. R. Soc. B* **372**, 20160161 (2017).
23. Levy, S. J., Kinsky, N. R., Mau, W., Sullivan, D. W. & Hasselmo, M. E. Hippocampal spatial memory representations in mice are heterogeneously stable. *Hippocampus* **31**, 244–260 (2021).
24. Kinsky, N. R. et al. Trajectory-modulated hippocampal neurons persist throughout memory-guided navigation. *Nat. Commun.* **11**, 2443 (2020).
25. Liberti, W. A., Perkins, L. N., Leman, D. P. & Gardner, T. J. An open source, wireless capable miniature microscope system. *J. Neural Eng.* **14**, 045001 (2017).
26. Genzel, D. & Yartsev, M. M. The fully automated bat (FAB) flight room: a human-free environment for studying navigation in flying bats and its initial application to the retrosplenial cortex. *J. Neurosci. Methods* **348**, 108970 (2021).
27. Dotson, N. M. & Yartsev, M. M. Nonlocal spatiotemporal representation in the hippocampus of freely flying bats. *Science* **373**, 242–247 (2021).
28. Sarel, A., Finkelstein, A., Las, L. & Ulanovsky, N. Vectorial representation of spatial goals in the hippocampus of bats. *Science*

**355**, 176–180 (2017).

29. Mallory, C. S. & Giocomo, L. M. Heterogeneity in hippocampal place coding. *Curr. Opin. Neurobiol.* **49**, 158–167 (2018).
30. McNaughton, B. L., Barnes, C. A. & O’Keefe, J. The contributions of position, direction, and velocity to single unit activity in the hippocampus of freely-moving rats. *Exp. Brain Res.* **52**, 41–49 (1983).
31. Wood, E. R., Dudchenko, P. A., Robitsek, R. J. & Eichenbaum, H. Hippocampal neurons encode information about different types of memory episodes occurring in the same location. *Neuron* **27**, 623–633 (2000).
32. Rubin, A., Yartsev, M. M. & Ulanovsky, N. Encoding of head direction by hippocampal place cells in bats. *J. Neurosci.* **34**, 1067–1080 (2014).
33. Geva-Sagiv, M., Romani, S., Las, L. & Ulanovsky, N. Hippocampal global remapping for different sensory modalities in flying bats. *Nat. Neurosci.* **19**, 952–958 (2016).
34. Kentros, C. G., Agnihotri, N. T., Streater, S., Hawkins, R. D. & Kandel, E. R. Increased attention to spatial context increases both place field stability and spatial memory. *Neuron* **42**, 283–295 (2004).
35. Muzzio, I. A. et al. Attention enhances the retrieval and stability of visuospatial and olfactory representations in the dorsal hippocampus. *PLoS Biol.* **7**, e1000140 (2009).
36. Dupret, D., O’Neill, J., Pleydell-Bouverie, B. & Csicsvari, J. The reorganization and reactivation of hippocampal maps predict spatial memory performance. *Nat. Neurosci.* **13**, 995–1002 (2010).
37. Tchernichovski, O., Benjamini, Y. & Golani, I. The dynamics of long-term exploration in the rat. Part I. A phase-plane analysis of the relationship between location and velocity. *Biol. Cybern.* **78**, 423–432 (1998).

38. Yartsev, M. M. The emperor's new wardrobe: rebalancing diversity of animal models in neuroscience research. *Science* **358**, 466–469 (2017).
39. Krakauer, J. W., Ghazanfar, A. A., Gomez-Marin, A., MacIver, M. A. & Poeppel, D. Neuroscience needs behavior: correcting a reductionist bias. *Neuron* **93**, 480–490 (2017).
40. Yovel, Y., Geva-Sagiv, M. & Ulanovsky, N. Click-based echolocation in bats: not so primitive after all. *J. Comp. Physiol. A* **197**, 515–530 (2011).
41. Holland, R. A., Waters, D. A. & Rayner, J. M. V. Echolocation signal structure in the Megachiropteran bat *Rousettus aegyptiacus* Geoffroy 1810. *J. Exp. Biol.* **207**, 4361–4369 (2004).
42. Lee, W.-J. et al. Tongue-driven sonar beam steering by a lingual-echolocating fruit bat. *PLoS Biol.* **15**, e2003148 (2017).
43. Danilovich, S. & Yovel, Y. Integrating vision and echolocation for navigation and perception in bats. *Sci. Adv.* **5**, eaaw6503 (2019).
44. El-Mansi, A. A., Al-Kahtani, M. A., Al-Sayyad, K. M., Ahmed, E. A. & Gad, A. R. Visual adaptability and retinal characterization of the Egyptian fruit bat (*Rousettus aegyptiacus*, Pteropodidae): new insights into photoreceptors spatial distribution and melanosomal activity. *Micron* **137**, 102897 (2020).
45. Ghosh, K. K. et al. Miniaturized integration of a fluorescence microscope. *Nat. Methods* **8**, 871–878 (2011).
46. Groot, A. D., Genderen, R. M. V., Coppens, J., Zeeuw, I. D. & Hoogland, T. M. NINscope: a versatile miniscope for multi-region circuit investigations. *eLife* **9**, e49987 (2019).
47. Shuman, T. et al. Breakdown of spatial coding and interneuron synchronization in epileptic mice. *Nat. Neurosci.* **23**, 229–238 (2020).

48. Cai, D. J. et al. A shared neural ensemble links distinct contextual memories encoded close in time. *Nature* **534**, 115–118 (2016).
49. Kügler, S., Kilic, E. & Bähr, M. Human synapsin 1 gene promoter confers highly neuron-specific long-term transgene expression from an adenoviral vector in the adult rat brain depending on the transduced area. *Gene Ther.* **10**, 337–347 (2003).
50. Barchi, J. R., Knowles, J. M. & Simmons, J. A. Spatial memory and stereotypy of flight paths by big brown bats in cluttered surroundings. *J. Exp. Biol.* **216**, 1053–1063 (2013).
51. Guizar-Sicairos, M., Thurman, S. T. & Fienup, J. R. Efficient subpixel image registration algorithms. *Opt. Lett.* **33**, 156–158 (2008).
52. Huang, L. et al. Relationship between simultaneously recorded spiking activity and fluorescence signal in GCaMP6 transgenic mice. *eLife* **10**, e51675 (2021).
53. Sheintuch, L. et al. Tracking the same neurons across multiple days in Ca<sup>2+</sup> imaging data. *Cell Rep.* **21**, 1102–1115 (2017).
54. Zhou, P. et al. Efficient and accurate extraction of in vivo calcium signals from microendoscopic video data. *eLife* **7**, e28728 (2018).
55. Pnevmatikakis, E. A. et al. Simultaneous denoising, deconvolution, and demixing of calcium imaging data. *Neuron* **89**, 285–299 (2015).
56. Skaggs, W. E., McNaughton, B. L., Wilson, M. A. & Gothard, K. M. An information-theoretic approach to deciphering the hippocampal code. In *Adv. Neural Inf. Process. Syst.* **5** (eds Hanson, S. J., Cowan, J. D. & Giles, C. L.) 1030–1037 (Morgan Kaufman, San Mateo, 1993).
57. Zhang, K., Ginzburg, I., McNaughton, B. L. & Sejnowski, T. J. Interpreting neuronal population activity by reconstruction: unified framework with application to hippocampal place cells. *J. Neurophysiol.* **79**, 1017–1044 (1998).

58. Carpenter, J., Blackstad, J., Dunn, B., Moser, E. I. & Moser, M.-B. Egocentric tuning in hippocampal neurons: using simulations to estimate misclassification rates. In *Proc. Soc. Neurosci.* P871.08 (2021).

## Acknowledgements

We thank the members of the Yartsev laboratory and D. Foster laboratory for discussion and comments; E. Azim, B. Styr, J. Widloski and K. Qi for reading and comments on the manuscript; P. Golshani; T. Shuman, D. Cai and G. Blair for initial training on calcium imaging surgical techniques; N. Doston and D. Genzel for assistance with the experimental set-up; V. Shvareva and A. Rakuljic for help with histology; H. Aaron, F. Ives and staff at the CRL Molecular Imaging Center for assistance with the use of core imaging facilities; Y. Minton and L. Loomis for experimental room maintenance and animal care; G. Lawson, K. Jensen, J. Frohlich and the staff of the Office of Laboratory Animal Care for support with animal husbandry and care. This research was supported by the New York Stem Cell Foundation (NYSCF-R-NI40), Air Force Office of Scientific Research (FA9550-17-1-0412), the Packard Fellowship (2017–66825), National Institute of Neurological Disorders and Stroke (R01NS118422-01), the Valle Foundation (VS-2020-34), The Office of Naval Research (N00014-21-1-2063), the Searle Scholars Program (SSP-2016-1412) (to M.M.Y.) and the National Science Foundation Graduate Research Fellowship (to T.A.S.).

## Author information

### Author notes

1. These authors contributed equally: William A. Liberti III, Tobias A. Schmid

## Affiliations

1. Department of Bioengineering, UC Berkeley, Berkeley, CA, USA

William A. Liberti III, Angelo Forli, Madeleine Snyder & Michael M. Yartsev

2. Helen Wills Neuroscience Institute, UC Berkeley, Berkeley, CA, USA

Tobias A. Schmid & Michael M. Yartsev

## Contributions

T.A.S., W.A.L. and M.M.Y. designed the research. W.A.L., T.A.S. and M.M.Y. analysed the data. T.A.S. performed experiments with help from W.A.L, A.F and M.S. Illustrations were made by M.S. The manuscript was written by W.A.L., T.A.S. and M.M.Y. with input from A.F. and M.S.

## Corresponding author

Correspondence to [Michael M. Yartsev](#).

## Ethics declarations

## Competing interests

The authors declare no competing interests.

## Peer review

## Peer review information

*Nature* thanks Roddy Grieves, Torkel Hafting and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

## Additional information

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## Extended data figures and tables

### Extended Data Fig. 1 The automated, human-free, flight room.

**a**, Panoramic perspective of the automated flight room showing the feeders, adjustable ceiling lights, foam walls and motion tracking cameras. **b**, View towards one of the interior walls of the flight room. Motion tracking camera is indicated. **c**, View of the feeders. Note that the ceiling lights (white squares) and tracking system are visible on the ceiling for all images. The ceiling lights were set to specific illumination levels during the flight experiments (Methods) and are shown here at maximum brightness level for visualization purposes.

### Extended Data Fig. 2 Imaged hippocampal neurons are spatially selective and collectively span the available environment.

**a**, 14 simultaneously recorded ROIs during a single free-foraging session from one bat. Red dots represent estimated spiking events superimposed on the flight trajectories (grey; Methods). The size of each dot indicates the relative size of the deconvolved spiking event per imaging frame (per 33 ms bin; Methods). **b**, **c**, Firing density is not increased around the virtual boundary and the spatial activity recorded across neurons spans the available environment. **b**, Example spatial firing density for one animal. The virtual boundary that bats had to cross in order to rebait reward feeders is indicated by a dashed line. **c**, Spatial firing density of deconvolved neural activity (that is, ‘spikes’) for all animals ( $n = 7$ ). Reward locations are indicated. Images in **b**, **c** are displayed from a top-down (XY) perspective. There was no significant difference in firing density within 1m of this boundary versus other locations in the room ( $P > 0.5$  for all bats, two-tailed Wilcoxon Rank Sum).

### Extended Data Fig. 3 Spatial behaviour and hippocampal neural tuning along structured flight paths.

**a, b**, Flights predominantly occur along a few highly repeated paths. **a**, After an initial period of exploration, a few common flight paths dominate the distribution of trajectories during the free-foraging session ( $n = 63$  sessions from 7 bats). Blue and black lines are the mean proportion of flights over time for the three most common flight paths and the unique flight paths, respectively. Shading is 95% confidence interval around the mean. 73% of all recorded flights (3,472/4,731 flights from 7 bats) occur along repeated flight paths, 54% are flown across the three most common routes (1,877/3,472 flights from 7 bats) and 70% are in the top five most common routes (2,429/3,472 flights from 7 bats). **b**, Percentage of the top three flight paths and all other flights in each session across all animals, with the mean and SEM indicated. Each point is the prevalence of the top three flight paths on each session in which they occur. Points are horizontally jittered to ease visualization. **c**, Percentage of significantly tuned ROIs for all bats ( $n = 7$ ; Methods). Each dot represents a single animal, the red line represents the mean for all seven bats, boxes are first and third quartiles and bars represent maximum and minimum across all bats. **d**, Percentage of ROIs with one or more fields per flight path (Methods).

[Source data](#)

#### [Extended Data Fig. 4 Neural tuning is distinct for different trajectories but is highly similar in overlapping portions.](#)

**a**, 2D projection (top-down view) of the mean flight path for the three most common flight paths for one bat. Note that the green and red trajectories are highly overlapping compared to the blue trajectory. **b**, Mean velocity profile for the same top three flight paths, with colours corresponding to those in **a**. **c**, Distribution of correlation scores (shown here as  $1-r$  for ease of comparison to Fig. 4) for all ROIs, relative to the red flight path. The red distribution is calculated by comparing even/odd flights of the red flight path shown in **a**. **d**, Same as **c** but only for the timeframe in which red and green flights overlap in space. **e**, Additional examples of overlapping flight paths. Overlapping paths are coloured black and non-overlapping paths are grey. Overlapping regions are coloured red. **f**, The black distribution is calculated by comparing neural tuning relative to even/odd flights that

comprise a flight path; the blue distribution is the comparison of neural tuning across distinct flight paths that partially overlap in some segments. The grey distribution compares neural tuning across flight paths that are similar in duration but do not overlap in space. **g**, Same as **f** but only comparing neural activity for the timeframe in which similar flight paths overlap in space. The red distribution is the comparison of neural tuning during the overlapping portions of distinct flight paths and the grey distribution is a time matched control for trajectories that do not overlap in space. For **f**, **g**, the black distribution is calculated by comparing neural tuning relative to even/odd flights that comprise a flight path; and red indicates a comparison of overlapping portions of distinct flight paths. **h**, **i**, Different flight paths can be accurately decoded using neural activity. **h**, Prediction accuracy for two bats (left and right confusion matrixes, respectively) where each class is a flight cluster. The percent accuracy and the number of classified flights is listed within each block (Methods). **i**, Decoder prediction accuracy for all bats and all flight paths with  $n > 10$  flights. Points are horizontally jittered for ease of visualization.

#### [Source data](#)

### Extended Data Fig. 5 Neural tuning is better explained by spatial rather than goal-vector tuning.

**a**, Example of the analysis approach. Shown is a 2D projection (top-down view) of the mean flight path for three flights of one bat. In this example, ROI distance and angle tuning for the grey flight path can be directly compared to a dissimilar flight path that terminates at the same goal location (in red, where shared angle/distance tuning is expected in the case of goal-vector tuning<sup>28</sup>) and to a similar flight path that ends in a different goal location (in blue, where goal-vector tuning is not expected, but shared angle/distance tuning might artefactually result from an overlap in spatial position). **b**, Distribution of correlation scores (shown here as  $1-r$  for ease of comparison to Fig. 4) for all ROIs binned by distance to goal (Methods) between pairs of similar flights (in blue) or pairs of dissimilar, goal terminating flight paths (in red). **c**. Same as **b** but binned by angle to the goal location. For **b**, **c**, neural tuning is significantly more similar across pairs of overlapping flight paths (distance:  $P = 1.4 \times 10^{-5}$  two-tailed

Wilcoxon Rank Sum; angle:  $P = 1.6 \times 10^{-5}$  two-tailed Wilcoxon rank sum); for dissimilar paths towards the same goal location, both goal and distance tuning distributions are not significantly different from what would be expected by chance (performing the same comparison after shuffling ROI identity,  $P > 0.05$  two-tailed Wilcoxon rank sum). The percent of ROIs that have the same significant tuning across two non-spatially overlapping flights to the same goal location is 4.1% (14/335 ROIs) for distance and 3.9% (13/335 ROIs) for angle (Methods). Therefore, we did not observe a pronounced goal-vector tuning independent of spatial position. This differs from a previous report in CA1 of flying bats<sup>28</sup> and could be due to increased false positive rates related to uncontrolled behavioural covariates, in agreement with a recent report<sup>58</sup>.

### Source data

### Extended Data Fig. 6 Tuning stability for neurons tracked over days.

**a**, The fraction of ROIs that can be confidently tracked, under ideal conditions (no experimenter refocusing or noticeably large change in field of view) as a function of days. Each dot represents the fraction of ROIs tracked for all bats. Tracking loss occurred at a rate of about 2% of ROIs per day (linear fit,  $R^2 = 0.67$ ,  $P = 3 \times 10^{-4}$ ). Dotted lines indicate 95% coefficient confidence intervals of the linear regression line in red. **b**, The cumulative distribution function (CDF) of unique ROIs that were included in the analysis, considering that some bats were tracked over longer intervals than others. **c**, Tracking contiguity showing the number of session gaps that exist between confidently tracked and aligned unique ROIs across days. **d**, Replotting the data in main text Fig. 2b, showing all underlying data points as a scatterplot. Error bars are 99% confidence intervals of the mean. Points are horizontally jittered for ease of visualization. Note that the number of ROI/flight-pair comparisons over days is decreasing ( $n = 914, 635, 490, 367, 208, 167, 117, 92, 45, 11$ ). **e**, Histogram of correlation values for all points in panel ‘d’ ( $n = 3,046$  possible ROI/flight path pairs). Note the lack of a pronounced second peak where tuning stability equals 0, as would be expected if a prominent subset of neurons were to lose or gain tuning. **f**, Scatterplot distribution of ROI correlations compared to day ‘1’

for the subset of consecutively tracked, flight path aligned ROIs ( $n = 360$  ROIs). Green box indicates the distribution median. **g**, Histogram of correlation values for all points in panel **f**. Note the lack of a pronounced second peak where tuning stability equals 0, as would be expected if a prominent subset of neurons were to lose or gain tuning. Points are horizontally jittered for ease of visualization. **h**, Distributions of the mean ROI peak times relative to the first day for all points in panel **f**, separated by day (colours). **i, j**, CA1 tuning persists despite a multi-day gap in flight path behaviour. **i**, The dominant three flight paths are coloured in red, green and blue. Note the five-day interval between sessions with green flight paths. **j**, Six example ROI timeseries aligned to the green flight path on day 1 (dark green) and day 6 (light green). Shading is the standard deviation of the mean fluorescent time series. Note the high similarity in the ROIs' timeseries that persists after a multi-day gap during which this flight path was not flown.

[Source data](#)

**Extended Data Fig. 7 Examples of abrupt, possibly biologically-driven, changes in neural participation over days.**

Maximum intensity projections of adjacent flight sessions that show several compelling examples of sharp in focus ROIs that could be unambiguously identified on one day but are clearly absent in the proceeding or following imaging session. Several of these examples are less likely to be a result of a slow change in imaging plane across days, which explains the majority of our tracking loss (that is, a slow change of ROI sharpness and intensity over consecutive sessions, as may be seen in some panels). Examples of abrupt changes in neural participation are sparse; panels **a–d** show different examples with panel **a** being the most obvious example found in our data set. Blue markers identify ROIs that are easily tracked on adjacent sessions. Red circles indicate an ROI that putatively ‘drops-out’ and green circles indicate an ROI that putatively ‘drops-in.’ Note that this phenomenon is uncommon and shown here to demonstrate the conservative approach: only 2.6% of clearly isolated ROIs identified on any given day show this effect (Methods).

## Extended Data Fig. 8 ROI tuning stability and flight consistency.

**a**, Flight-to-flight consistency varies along the phase of flight. Displayed are all ROIs tuned to the flight path shown in Fig. 3a, b. Each point represents a significantly tuned ROI's stability (that is, one minus the normalized coefficient of variation (CV)) at the flight phase it is tuned to (determined by the peak of the mean flight-aligned deconvolved activity trace). Red dotted lines indicate takeoff and landing times. The solid blue line is the running average of all points. **b**, Flight variability as a function of flight phase, sampled on the same interval as in **a**. Flight variability was highest in the middle of the flight and lowest during the takeoff and landing. The left y-axis units are the mean Euclidean distance of all individual flights from the centre of the mean flight path. The blue dotted line is the same as the blue solid line from **a**, normalized from 0-1 (right y-axis). **c**, More structured behaviour in freely flying bats results in the estimation of a more stable CA1 hippocampal responses over days. Longitudinal stability of hippocampal neural responses when considering either the most common structured flight path (that is, the flight path comprised of the largest number of flights, in red) and all non-repeated (that is, unstructured) flights in the same session (in blue) for  $n = 7$  bats. Correlation is computed for 2D rate-maps relative to the first day of the experiment. Thin lines are different bats/flight paths and thick lines are the average across animals. Error bars represent 95% confidence intervals of the mean.

[Source data](#)

## Extended Data Fig. 9 Perceived sensory-based remapping when considering all flight trajectories before controlling for flight behaviour variability across lights-on and lights-off conditions.

Shown is the distribution of 2D rate map correlation scores for all ROIs when considering all flight paths between either the two lights-on sessions (teal) or lights-on versus lights-off conditions (grey). These distributions are significantly different ( $n = 178$  ROIs,  $P = 7.04 \times 10^{-55}$ , two-tailed Wilcoxon rank sum test). \*\*\*  $P < 0.0001$ .

## Source data

### Extended Data Fig. 10 Similarities and differences in flight behaviour between the lights-on vs. lights-off cue conditions.

**a**, Aggregated flight paths that occurred across 10 consecutive days of experiments for one example bat, divided into successive lights-on, lights-off and lights-on' periods (ordered from left to right). Coloured are the four most common flight paths. The red and blue flight paths occurred predominantly in lights-on and lights-on' sessions. The green and magenta paths occurred predominantly in the lights-off condition. **b**, Distribution of the occurrence of the four most prevalent flight paths, coloured according to **a**, during different phases of the experiment across lights-on and lights-off sessions. Flight preference changed dramatically during each period. **c**, Aggregated flight paths across ten consecutive days for one representative bat divided into chronological thirds of each session where light levels were held constant. Shown are the three most common flight paths in red, blue and green. **d**, Distribution of the occurrence of the four most prevalent flight paths (colours) during different phases of the experiment across each third of ten sessions. **e**, Proportion of unique (that is, unstructured) flight paths increase on average in the lights-off cue condition (dark) compared to when the lights are kept on for the duration of the session. The blue line indicates the mean relative proportion of unique flights flown in the lights-on/off experiment. The dark grey line is the mean proportion of unstructured flights for bats where the light is kept on for the duration of the experiment. Shading for both plots represent 95% confidence intervals of the mean. **f**, The mean number of flights per minute is not significantly different in the light or dark conditions ( $P > 0.05$  two-tailed Wilcoxon rank sum test,  $n = 80$  and 40 binned minutes in the lights-on and lights-off periods respectively). Points are horizontally jittered for ease of visualization. **g**, Flight duration as function of cue condition. Flight duration is significantly longer in the lights-off condition, ( $P = 0.0038$ , two-tailed Wilcoxon rank sum test) although the effect size is small (that is, the distributions are highly overlapping;  $n = 1,567$  and 746 flights in the lights-on and lights-off periods respectively). Points are horizontally jittered for ease of visualization. **h, i**, Across-day and across-condition behavioural stability for the lights-on versus lights-off experiments. **h**, Average

correlation of repeated paths relative to day one for each bat ( $n = 5$  flight paths). Each colour indicates a different flight path. Bars indicate 95% confidence intervals of the mean correlation on each day. **i**, Correlation values of shared flights to the mean of their flight path in the first lights-on period. Flights along the same path were not significantly different between the light and dark cue condition ( $P = 0.21$  two-tailed Wilcoxon rank sum test,  $n = 268, 150, 267$  flights for lights-on, lights-off and lights-on conditions, respectively). Points are horizontally jittered for ease of visualization. \*  $P < 0.01$ .

## [Source data](#)

## **Supplementary information**

### [Reporting Summary](#)

### **Source data**

#### [Source Data Fig. 3](#)

#### [Source Data Fig. 4](#)

#### [Source Data Extended Data Fig. 3](#)

#### [Source Data Extended Data Fig. 4](#)

#### [Source Data Extended Data Fig. 5](#)

#### [Source Data Extended Data Fig. 6](#)

#### [Source Data Extended Data Fig. 8](#)

#### [Source Data Extended Data Fig. 9](#)

#### [Source Data Extended Data Fig. 10](#)

# Rights and permissions

[Reprints and Permissions](#)

## About this article

### Cite this article

Liberti, W.A., Schmid, T.A., Forli, A. *et al.* A stable hippocampal code in freely flying bats. *Nature* **604**, 98–103 (2022).  
<https://doi.org/10.1038/s41586-022-04560-0>

- Received: 16 June 2021
- Accepted: 17 February 2022
- Published: 30 March 2022
- Issue Date: 07 April 2022
- DOI: <https://doi.org/10.1038/s41586-022-04560-0>

### Share this article

Anyone you share the following link with will be able to read this content:

[Get shareable link](#)

Sorry, a shareable link is not currently available for this article.

[Copy to clipboard](#)

Provided by the Springer Nature SharedIt content-sharing initiative

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

- Article
- [Published: 30 March 2022](#)

# Entropy of city street networks linked to future spatial navigation ability

- [A. Coutrot](#) [ORCID: orcid.org/0000-0001-9569-3548<sup>1</sup>](#),
- [E. Manley<sup>2,3</sup>](#),
- [S. Goodroe](#) [ORCID: orcid.org/0000-0003-1468-8304<sup>4</sup>](#),
- [C. Gahnstrom](#) [ORCID: orcid.org/0000-0002-9989-5136<sup>4</sup>](#),
- [G. Filomena](#) [ORCID: orcid.org/0000-0001-6380-7264<sup>5</sup>](#),
- [D. Yesiltepe<sup>6</sup>](#),
- [R. C. Dalton<sup>6</sup>](#),
- [J. M. Wiener<sup>7</sup>](#),
- [C. Hölscher<sup>8</sup>](#),
- [M. Hornberger<sup>9</sup>](#) &
- [H. J. Spiers](#) [ORCID: orcid.org/0000-0002-6792-6356<sup>4</sup>](#)

[Nature](#) volume 604, pages 104–110 (2022)

- 8879 Accesses
- 1071 Altmetric
- [Metrics details](#)

## Subjects

- [Human behaviour](#)
- [Intelligence](#)

# Abstract

The cultural and geographical properties of the environment have been shown to deeply influence cognition and mental health<sup>1,2,3,4,5,6</sup>. Living near green spaces has been found to be strongly beneficial<sup>7,8,9,10,11</sup>, and urban residence has been associated with a higher risk of some psychiatric disorders<sup>12,13,14</sup>—although some studies suggest that dense socioeconomic networks found in larger cities provide a buffer against depression<sup>15</sup>. However, how the environment in which one grew up affects later cognitive abilities remains poorly understood. Here we used a cognitive task embedded in a video game<sup>16</sup> to measure non-verbal spatial navigation ability in 397,162 people from 38 countries across the world. Overall, we found that people who grew up outside cities were better at navigation. More specifically, people were better at navigating in environments that were topologically similar to where they grew up. Growing up in cities with a low street network entropy (for example, Chicago) led to better results at video game levels with a regular layout, whereas growing up outside cities or in cities with a higher street network entropy (for example, Prague) led to better results at more entropic video game levels. This provides evidence of the effect of the environment on human cognition on a global scale, and highlights the importance of urban design in human cognition and brain function.

This is a preview of subscription content

## Access options

Subscribe to Nature+

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

\$29.99

monthly

[Subscribe](#)

[Subscribe to Journal](#)

Get full journal access for 1 year

\$199.00

only \$3.90 per issue

[Subscribe](#)

All prices are NET prices.

VAT will be added later in the checkout.

Tax calculation will be finalised during checkout.

[Buy article](#)

Get time limited or full article access on ReadCube.

\$32.00

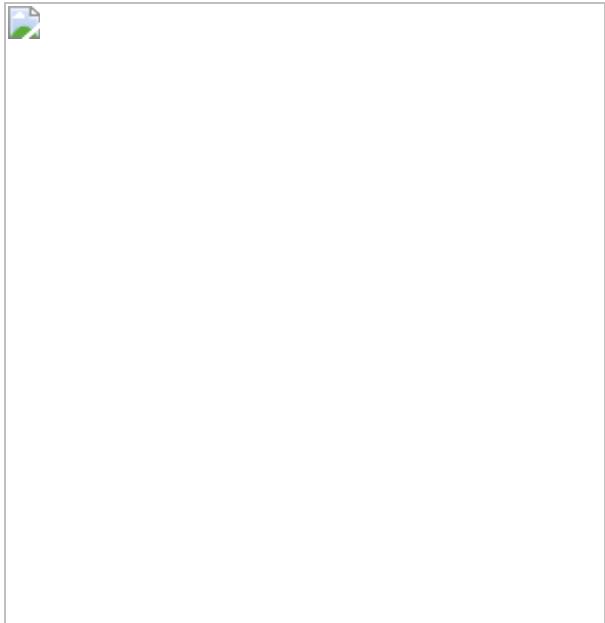
[Buy](#)

All prices are NET prices.

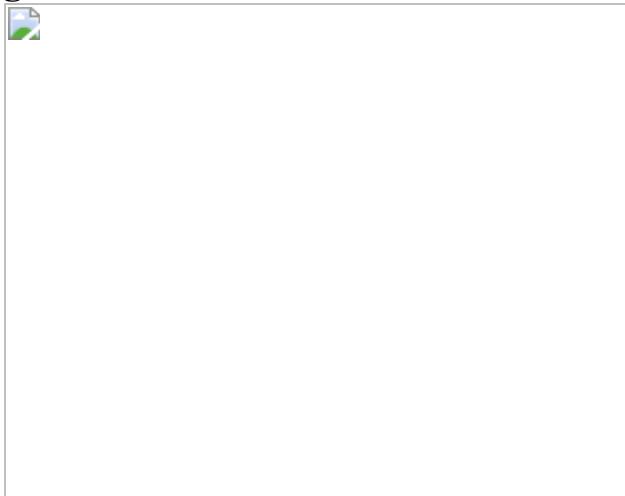
### **Additional access options:**

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

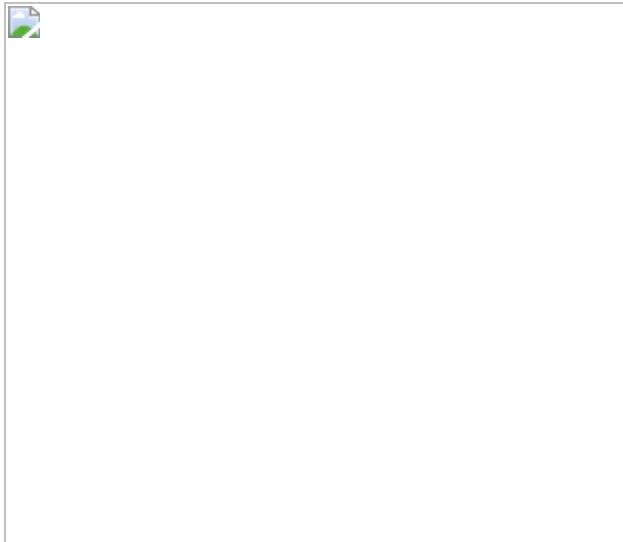
**Fig. 1: Wayfinding task.**



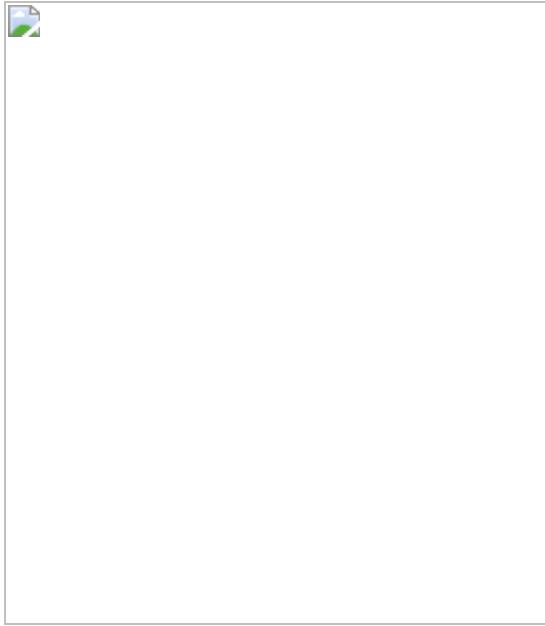
**Fig. 2: SNE and environment effect in 38 countries.**



**Fig. 3: Comparison of SNE with other measures of city complexity.**



**Fig. 4: Participants are accurate at navigating more entropic game levels when they grew up in more entropic environments.**



## Data availability

A dataset with the preprocessed trajectory lengths and demographic information is available at [https://osf.io/7nqw6/?view\\_only=6af022f2a7064d4d8a7e586913a1f157](https://osf.io/7nqw6/?view_only=6af022f2a7064d4d8a7e586913a1f157). Owing to its considerable size (around 1 terabyte), the dataset with the full trajectories is available on a dedicated server: <https://shqdata.z6.web.core.windows.net/>. We also set up a portal where researchers can invite a targeted group of

participants to play SHQ and generate data about their spatial navigation capabilities. Those invited to play the game will be sent a unique participant key, generated by the SHQ system according to the criteria and requirements of a specific project.

<https://seaheroquest.alzheimersresearchuk.org/> Access to the portal will be granted for non-commercial purposes. Future publications based on this dataset should add ‘Sea Hero Quest Project’ as a co-author.

## Code availability

The Python and MATLAB (R2018a) code that allows the presented analyses to be reproduced is available along the preprocessed trajectory lengths and demographic information at [https://osf.io/7nqw6/?view\\_only=6af022f2a7064d4d8a7e586913a1f157](https://osf.io/7nqw6/?view_only=6af022f2a7064d4d8a7e586913a1f157).

## References

1. Kempermann, G., Kuhn, H. G. & Gage, F. H. More hippocampal neurons in adult mice living in an enriched environment. *Nature* **386**, 493–495 (1997).
2. Hackman, D. A., Farah, M. J. & Meaney, M. J. Socioeconomic status and the brain: mechanistic insights from human and animal research. *Nat. Rev. Neurosci.* **11**, 651–659 (2010).
3. May, A. Experience-dependent structural plasticity in the adult human brain. *Trends Cogn. Sci.* **15**, 475–482 (2011).
4. Van Praag, H., Kempermann, G. & Gage, F. H. Neural consequences of environmental enrichment. *Nat. Rev. Neurosci.* **1**, 191–198 (2000).
5. Freund, J. et al. Emergence of individuality in genetically identical mice. *Science* **340**, 756–759 (2013).
6. Clemenson, G. D., Deng, W. & Gage, F. H. Environmental enrichment and neurogenesis: from mice to humans. *Curr. Opin. Behav. Sci.* **4**, 56–62 (2015).

7. Kardan, O. et al. Neighborhood greenspace and health in a large urban center. *Sci. Rep.* **5**, 11610 (2015).
8. Dadvand, P. et al. Green spaces and cognitive development in primary schoolchildren. *Proc. Natl Acad. Sci. USA* **112**, 7937–7942 (2015).
9. Engemann, K. et al. Residential green space in childhood is associated with lower risk of psychiatric disorders from adolescence into adulthood. *Proc. Natl Acad. Sci. USA* **116**, 5188–5193 (2019).
10. Berman, M. G., Stier, A. J. & Akcelik, G. N. Environmental neuroscience. *Am. Psychol.* **74**, 1039–1052 (2019).
11. Bratman, G. N. et al. Nature and mental health: an ecosystem service perspective. *Sci. Adv.* **5**, eaax0903 (2019).
12. Lederbogen, F. et al. City living and urban upbringing affect neural social stress processing in humans. *Nature* **474**, 498–501 (2011).
13. Kühn, S. et al. In search of features that constitute an “enriched environment” in humans: Associations between geographical properties and brain structure. *Sci. Rep.* **7**, 11920 (2017).
14. Carey, I. M. et al. Are noise and air pollution related to the incidence of dementia? A cohort study in London, England. *BMJ Open* **8**, e022404 (2018).
15. Stier, A. et al. Rethinking depression in cities: evidence and theory for lower rates in larger urban areas. Preprint at <https://doi.org/10.1101/2020.08.20.20179036> (2020).
16. Coutrot, A. et al. Global determinants of navigation ability. *Curr. Biol.* **28**, 2861–2866 (2018).
17. Malanchini, M. et al. Evidence for a unitary structure of spatial cognition beyond general intelligence. *npj Sci. Learn.* **5**, 9 (2020).
18. Spiers, H. J. & Maguire, E. A. Thoughts, behaviour, and brain dynamics during navigation in the real world. *Neuroimage* **31**, 1826–

1840 (2006).

19. Maguire, E. A., Woollett, K. & Spiers, H. J. London taxi drivers and bus drivers: a structural MRI and neuropsychological analysis. *Hippocampus* **16**, 1091–1101 (2006).
20. Xu, J. et al. Global urbanicity is associated with brain and behaviour in young people. *Nat. Hum. Behav.* <https://doi.org/10.1038/s41562-021-01204-7> (2021).
21. Coutrot, A. et al. Virtual navigation tested on a mobile app is predictive of real-world wayfinding navigation performance. *PLoS ONE* **14**, e0213272 (2019).
22. Spiers, H. J., Coutrot, A. & Hornberger, M. Explaining world-wide variation in navigation ability from millions of people: citizen science project Sea Hero Quest. *Top. Cogn. Sci.* <https://doi.org/10.1111/tops.12590> (2021).
23. Sutherland, R. J. & Hamilton, D. A. Rodent spatial navigation: at the crossroads of cognition and movement. *Neurosci. Biobehav. Rev.* **28**, 687–697 (2004).
24. Epstein, R. A., Patai, E. Z., Julian, J. B. & Spiers, H. J. The cognitive map in humans: spatial navigation and beyond. *Nat. Neurosci.* **20**, 1504 (2017).
25. Boeing, G. OSMnx: new methods for acquiring, constructing, analyzing, and visualizing complex street networks. *Comput. Environ. Urban Syst.* **65**, 126–139 (2017).
26. Coughlan, G. et al. Toward personalized cognitive diagnostics of at-genetic-risk Alzheimer’s disease. *Proc. Natl Acad. Sci. USA* **116**, 9285–9292 (2019).
27. Klencklen, G., Després, O. & Dufour, A. What do we know about aging and spatial cognition? Reviews and perspectives. *Ageing Res. Rev.* **11**, 123–135 (2012).

28. Lester, A. W., Moffat, S. D., Wiener, J. M., Barnes, C. A. & Wolbers, T. The aging navigational system. *Neuron* **95**, 1019–1035 (2017).
29. Nazareth, A., Huang, X., Voyer, D. & Newcombe, N. A meta-analysis of sex differences in human navigation skills. *Psychon. Bull. Rev.* **26**, 1503–1528 (2019).
30. Ritchie, S. J. & Tucker-Drob, E. M. How much does education improve intelligence? A meta-analysis. *Psychol. Sci.* **29**, 1358–1369 (2018).
31. Ulrich, S., Grill, E. & Flanigin, V. L. Who gets lost and why: a representative cross-sectional survey on sociodemographic and vestibular determinants of wayfinding strategies. *PLoS ONE* **14**, e0204781 (2019).
32. Fuchs, F. et al. Exposure to an enriched environment up to middle age allows preservation of spatial memory capabilities in old age. *Behav. Brain Res.* **299**, 1–5 (2016).
33. Lynch, K. *The Image of the City* (The MIT Press, 1960).
34. Marshall, S. *Streets and Patterns* (Spon Press, 2005).
35. Watts, A., Ferdous, F., Diaz Moore, K. & Burns, J. M. Neighborhood integration and connectivity predict cognitive performance and decline. *Gerontol. Geriatr. Med.*  
<https://doi.org/10.1177/2333721415599141> (2015).
36. Koohsari, M. J. et al. Cognitive function of elderly persons in Japanese neighborhoods: the role of street layout. *Am. J. Alzheimers Dis. Other Demen.* **34**, 381–389 (2019).
37. Bongiorno, C. et al. Vector-based pedestrian navigation in cities. *Nat. Comput. Sci.* **1**, 678–685 (2021).
38. Boeing, G. A multi-scale analysis of 27,000 urban street networks: every US city, town, urbanized area, and Zillow neighborhood.

*Environ. Plann. B Urban Anal. City Sci.* **47**, 590–608 (2018).

39. Shannon, C. E. A mathematical theory of communication. *Bell Syst. Tech. J.* **27**, 379–423 (1948).
40. Barthélemy, M. Spatial networks. *Phys. Rep.* **499**, 1–101 (2011).
41. Gudmundsson, A. & Mohajeri, N. Entropy and order in urban street networks. *Sci. Rep.* **3**, 3324 (2013).
42. Batty, M., Morphet, R., Masucci, P. & Stanilov, K. Entropy, complexity, and spatial information. *J. Geogr. Syst.* **16**, 363–385 (2014).
43. Boeing, G. Urban spatial order: street network orientation, configuration, and entropy. *Appl. Netw. Sci.* **67**, 1–20 (2019).
44. McNamee, D., Wolpert, D. & Lengyel, M. Efficient state-space modularization for planning: theory, behavioral and neural signatures. In *Advances in Neural Information Processing Systems 29 (NIPS 2016)* (eds Lee, D. et al.) 4511–4519 (Curran Associates, 2016).
45. Wiener, J. M., Schnee, A. & Mallot, H. A. Use and interaction of navigation strategies in regionalized environments. *J. Env. Psychol.* **24**, 475–493 (2004).
46. Brunyé, T. T. et al. Strategies for selecting routes through real-world environments: Relative topography, initial route straightness, and cardinal direction. *PLoS ONE* **10**, e0124404 (2015).
47. Ekstrom, A. D., Spiers, H. J., Bohbot, V. D. & Rosenbaum, R. S. *Human Spatial Navigation* (Princeton University Press, 2018).
48. Salon, D. Heterogeneity in the relationship between the built environment and driving: focus on neighborhood type and travel purpose. *Res. Transp. Econ.* **52**, 34–45 (2015).
49. Lenormand, M., Bassolas, A. & Ramasco, J. J. Systematic comparison of trip distribution laws and models. *J. Transp. Geogr.* **51**, 158–169

(2016).

50. Nazareth, A., Weisberg, S. M., Margulis, K. & Newcombe, N. S. Charting the development of cognitive mapping. *J. Exp. Child Psychol.* **170**, 86–106 (2018).
51. Montello, D. R. A conceptual model of the cognitive processing of environmental distance information. In *Spatial Information Theory: 9th International Conference, COSIT 2009* (eds. Hornsby, K. S. et al.) 1–17 (Springer, 2009).
52. Masucci, A. P., Arcaute, E., Hatna, E., Stanilov, K. & Batty, M. On the problem of boundaries and scaling for urban street networks. *J. R. Soc. Interface* **12**, 20150763 (2015).
53. Giacomin, D. J. & Levinson, D. M. Road network circuitry in metropolitan areas. *Environ. Plann. B Plann. Des.* **42**, 1040–1053 (2015).
54. Jiang, B. & Claramunt, C. Topological analysis of urban street networks. *Environ. Plann. B Plann. Des.* **31**, 151–162 (2004).
55. Porta, S. et al. Street centrality and densities of retail and services in Bologna, Italy. *Environ. Plann. B Plann. Des.* **36**, 450–465 (2009).
56. Javadi, A.-H. et al. Hippocampal and prefrontal processing of network topology to simulate the future. *Nat. Commun.* **8**, 14652 (2017).
57. Jiang, B. & Claramunt, C. A structural approach to the model generalization of an urban street network. *GeoInformatica* **8**, 157–171 (2004).
58. Filomena, G., Verstegen, J. A. & Manley, E. A computational approach to ‘The Image of the City’. *Cities* **89**, 14–25 (2019).
59. Mou, W., McNamara, T. P., Valiquette, C. M. & Rump, B. Allocentric and egocentric updating of spatial memories. *J. Exp. Psychol. Learn. Mem. Cogn.* **30**, 142 (2004).

60. Tversky, B. Distortions in memory for maps. *Cogn. Psychol.* **13**, 407–433 (1981).
61. Sadalla, E. K. & Magel, S. G. The perception of traversed distance. *Environ. Behav.* **12**, 65–79 (1980).
62. Spiers, H. J. & Maguire, E. A. A navigational guidance system in the human brain. *Hippocampus* **17**, 618–626 (2007).
63. Howard, L. R. et al. The hippocampus and entorhinal cortex encode the path and Euclidean distances to goals during navigation. *Curr. Biol.* **24**, 1331–1340 (2014).
64. Spiers, H. J. & Barry, C. Neural systems supporting navigation. *Curr. Opin. Behav. Sci.* **1**, 47–55 (2015).
65. Douglas, D. H. & Peucker, T. K. Algorithms for the reduction of the number of points required to represent a digitized line or its caricature. *Cartographica* **10**, 112–122 (1973).
66. Hentschke, H. & Stützgen, M. C. Computation of measures of effect size for neuroscience data sets. *Eur. J. Neurosci.* **34**, 1887–1894 (2011).

## Acknowledgements

This research is part of the Sea Hero Quest initiative funded and supported by Deutsche Telekom. Alzheimer’s Research UK (ARUK-DT2016-1) funded the analysis; Glitchers designed and produced the game; and Saatchi and Saatchi London managed the creation of the game. The geographical information used in this study has been made available by OSM contributors under the Open Database License (<https://www.openstreetmap.org/copyright>).

## Author information

### Affiliations

1. LIRIS, CNRS, University of Lyon, Lyon, France  
A. Coutrot
2. Centre for Advanced Spatial Analysis, University College London, London, UK  
E. Manley
3. School of Geography, University of Leeds, Leeds, UK  
E. Manley
4. Institute of Behavioural Neuroscience, Department of Experimental Psychology, Division of Psychology and Language Sciences, University College London, London, UK  
S. Goodroe, C. Gahnstrom & H. J. Spiers
5. Institute for Geoinformatics, University of Münster, Münster, Germany  
G. Filomena
6. Department of Architecture and Built Environment, Northumbria University, Newcastle upon Tyne, UK  
D. Yesiltepe & R. C. Dalton
7. Department of Psychology, Ageing and Dementia Research Centre, Bournemouth University, Poole, UK  
J. M. Wiener
8. ETH Zürich, Swiss Federal Institute of Technology, Zürich, Switzerland  
C. Hölscher
9. Norwich Medical School, University of East Anglia, Norwich, UK

M. Hornberger

## Contributions

H.J.S., M.H. and A.C. supervised the project. H.J.S., M.H., A.C., S.G., C.G., R.C.D., J.M.W. and C.H. designed the research. A.C., E.M., G.F. and D.Y. analysed data. A.C., E.M. and H.J.S. wrote the paper.

## Corresponding authors

Correspondence to [A. Coutrot](#) or [H. J. Spiers](#).

## Ethics declarations

## Competing interests

The authors declare no competing interests.

## Peer review

### Peer review information

*Nature* thanks Marc Berman, Luis Bettencourt, Mary Hegarty, Nora Newcombe and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

## Additional information

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## Extended data figures and tables

[Extended Data Fig. 1 Colour-coded world maps.](#)

**a**, Sample size. **b**, Proportion of city participants. **c**, Environment effect size computed from a LMM predicting wayfinding performance, with fixed effects for age, gender and education, and random environment slopes clustered by country. The environment effect sizes are the environment slopes clustered by country, identical to the values in Fig. [2a](#).

### [Extended Data Fig. 2 Association between age, home environment, country and path integration performance.](#)

**a**, Path integration performance as a function of age for male and female participants who grew up in city and non-city environments. Path integration performance is averaged within 5-year windows, centre values correspond to the means. **b**, Difference of the effect of growing up outside cities on path integration performance across countries. We fit a logistic mixed model for path integration performance, with fixed effects for age, gender and education, and random environment slopes clustered by country, see [Supplementary Methods](#). Positive values indicate an advantage for participants raised outside cities. **c**, SNE as a function of the environment effect size (random environment slope) in each country, as in Fig. [2b](#), see [Supplementary Methods](#). All error bars correspond to standard errors,  $n = 182,122$  participants.

### [Extended Data Fig. 3 Environment effect size across age, gender and level of education.](#)

Effect size is quantified with Hedge's  $g$ , within five-year windows. Positive values correspond to an advantage for participants who grew-up outside cities. Error bars correspond to 95% CI and the centre values correspond to the means.

### [Extended Data Fig. 4 Wayfinding performance in city and non-city environments across age, in each country.](#)

Wayfinding performance is averaged within 10-year windows. Error bars correspond to standard errors and centre values correspond to the means. Note that these values correspond to raw wayfinding performance; that is,

they have not been corrected for age, gender or education. Note: Vietnam and Albania  $y$  axis lower bound is 0 to allow display of data points, instead of 0.5 for the rest of the countries. Altogether, we included  $n = 397,162$  participants.

### **Extended Data Fig. 5 Examples of city street networks.**

**a**, The road networks of New York City (USA, right) and London (UK, left) have been partitioned using the Louvain community detection algorithm on the dual graph, setting edge cost as angular change. The road networks within a  $3 \times 3 \text{ km}^2$  box around the city centres are represented. **b**, Street network of the 10 biggest cities in terms of population in Argentina and in Romania. We used OSMnx to gather the “drive” OSM network within  $1,000 \times 1,000\text{-m}^2$  boxes around each city centre. The reasons behind these differences are mostly historical. In South America, grid city design is characteristic of Hispanic American colonization, while disorganized street networks correspond to the typical organic street pattern of old European city cores.

### **Extended Data Fig. 6 Association between GDP per capita and SNE in City Hero Quest, a city-themed version of Sea Hero Quest.**

**a**, GDP per capita as a function of SNE. **b**, Screenshot from SHQ (left) and CHQ (right). **c**, Subset of SHQ levels used in the second experiment run on Prolific. **d**, CHQ levels used in the second experiment.

### **Extended Data Fig. 7 Association between age, home environment, country and wayfinding performance.**

**a**, Wayfinding performance as a function of age for participants who grew up in city, suburb, mixed and rural environments. Data points correspond to the wayfinding performance averaged within 5-year windows. **b**, Difference in the effect of growing up outside cities on wayfinding performance across countries. We fit a linear mixed model for wayfinding performance, with fixed effects for age, gender and education, and random environment slopes

clustered by country, as in Fig. 2a. Suburbs, mixed and rural environment slopes are represented, with City environment as baseline. Positive values correspond to an advantage compared to growing up in cities. Countries are ranked according to their suburb slope. The slopes of the different non-city environments are highly correlated: Pearson's  $r(\text{suburb, mixed}) = 0.97, p < 0.001$ ,  $r(\text{suburb, rural}) = 0.72, p < 0.001$ ,  $r(\text{mixed, rural}) = 0.53, p < 0.001$ . The country ranking is very similar to the one with only 2 classes (city / non-city): Spearman's  $r(\text{non-city, suburb}) = 0.85, p < 0.001$ ,  $r(\text{non-city, mixed}) = 0.73, p < 0.001$ ,  $r(\text{non-city, rural}) = 0.94, p < 0.001$ .  $P$  values are from a t-test testing the hypothesis of no correlation against the alternative hypothesis of a nonzero correlation. c, Pairwise differences between random environment slopes shown in panel b, averaged over countries. We show that the average difference in effect size between the city environment and the other 3 environments (city-rural, city-mixed, city-suburb) are around 10 times larger than the difference between the 'non-city' environments (rural-mixed, mixed-suburb, rural-suburb). This supports the approach to cluster together rural, mixed and suburb environments. All error bars correspond to standard errors,  $n = 397,162$  participants.

### Extended Data Fig. 8 Environment effect size and city complexity measures in high-SNE and low-SNE countries.

In each of the 380 included cities we computed a range of metrics to quantify different aspects of its complexity. We then took an average of these metrics weighted by the city population to have one value per country. We normalized these values by dividing them by their maximum. Network-based metrics - On top of the SNE used in this study, we computed other graph-theoretic measures commonly considered for spatial analysis of cities: average street length, circuity, neighbourhood degree, clustering coefficient, closeness centrality, betweenness centrality, and degree centrality. Route-based metrics - we simulated 1,000 routes in each city, and quantified five key variables derived from each route: number of unique streets, number of transitions in the partitions in street network structure, deviation from regular  $90^\circ$  turns at each turn, overall deviation from the target and number of turns above  $50^\circ$ . Individual data points correspond to countries ( $n = 38$ ). In the box plots, the horizontal bar represents the sample median, the hinges represent the first and third quartiles, and the whiskers

extend from the hinges to the largest/lowest value no further than  $\pm 1.5 \times$  IQR from the hinge (where IQR is the inter-quartile range).

### Extended Data Fig. 9 SNE across reported home environments.

SNE computed at the home addresses of the 599 participants to the follow-up experiment CHQ as a function of the reported type of home environment. The size of the square boxes used to compute the SNE were adjusted for the average street density within each reported environment (see [Supplementary Methods](#)). Error bars correspond to standard errors and centre values correspond to the means.

### Extended Data Fig. 10 Estimation of the robustness of the Pearson's correlation between SNE and environment effect size.

Bootstrapped correlation coefficients computed from 1,000 resampling with replacement. **a**, Histogram of the computed correlation coefficients. We obtained  $r = -0.60$ , 95% CI =  $[-0.78 \text{ } -0.30]$ . **b**, Regression lines for each sample. **c**, Pearson's correlation between environment effect size and different SNE calculations. The SNE set in bold is the one used in this manuscript. OSM = OpenStreetMaps, GM = Google Maps.  $P$  values are from a t-test testing the hypothesis of no correlation against the alternative hypothesis of a nonzero correlation.

## Supplementary information

### Supplementary Information

This file contains Supplementary Discussion; Supplementary Notes; Supplementary Methods and Supplementary References

### Reporting Summary

### Supplementary Table 1

For each country, the number of participants included in the analysis ( $N$ ), their mean age and standard deviation, the proportion of male versus female individuals, tertiary versus secondary education, city versus non-city home environment.

## **Supplementary Table 2**

For each country, the 10 biggest cities in term of population with their route-based metrics (unique streets, crossed partitions, turns above  $50^\circ$ , snap deviation, target deviation) and their network-based metrics (SNE, street length, circuity, neighbourhood degree, clustering coefficient, closeness centrality, betweenness centrality, degree centrality).

## **Peer Review File**

## **Supplementary Video 1**

Examples of navigation in two Sea Hero Quest levels: level 27 (left) and level 58 (right).

## **Supplementary Video 2**

Example of navigation in one City Hero Quest level.

## **Rights and permissions**

### **Reprints and Permissions**

## **About this article**

### **Cite this article**

Coutrot, A., Manley, E., Goodroe, S. *et al.* Entropy of city street networks linked to future spatial navigation ability. *Nature* **604**, 104–110 (2022).  
<https://doi.org/10.1038/s41586-022-04486-7>

- Received: 08 July 2020
- Accepted: 28 January 2022
- Published: 30 March 2022
- Issue Date: 07 April 2022
- DOI: <https://doi.org/10.1038/s41586-022-04486-7>

## Share this article

Anyone you share the following link with will be able to read this content:

[Get shareable link](#)

Sorry, a shareable link is not currently available for this article.

[Copy to clipboard](#)

Provided by the Springer Nature SharedIt content-sharing initiative

## [Winding roads could make you a better navigator](#)

- Shamini Bundell
- Nick Petrić Howe

Nature Podcast 30 Mar 2022

---

This article was downloaded by **calibre** from <https://www.nature.com/articles/s41586-022-04486-7>

- Article
- [Published: 30 March 2022](#)

# Human distal lung maps and lineage hierarchies reveal a bipotent progenitor

- [Preetish Kadur Lakshminarasimha Murthy<sup>1,na1</sup>](#),
- [Vishwaraj Sontake](#) ORCID: [orcid.org/0000-0002-8532-7680<sup>1,na1</sup>](https://orcid.org/0000-0002-8532-7680),
- [Aleksandra Tata<sup>1</sup>](#),
- [Yoshihiko Kobayashi<sup>1,16 nAff14</sup>](#),
- [Lauren Macadlo<sup>1</sup>](#),
- [Kenichi Okuda](#) ORCID: [orcid.org/0000-0001-9341-2730<sup>2</sup>](https://orcid.org/0000-0001-9341-2730),
- [Ansley S. Conchola](#) ORCID: [orcid.org/0000-0002-2385-6712<sup>3</sup>](https://orcid.org/0000-0002-2385-6712),
- [Satoko Nakano<sup>2</sup>](#),
- [Simon Gregory<sup>4</sup>](#),
- [Lisa A. Miller<sup>5,6</sup>](#),
- [Jason R. Spence](#) ORCID: [orcid.org/0000-0001-7869-3992<sup>7</sup>](https://orcid.org/0000-0001-7869-3992),
- [John F. Engelhardt](#) ORCID: [orcid.org/0000-0003-2389-9277<sup>8,9</sup>](https://orcid.org/0000-0003-2389-9277),
- [Richard C. Boucher<sup>2</sup>](#),
- [Jason R. Rock<sup>10</sup>](#),
- [Scott H. Randell<sup>11</sup>](#) &
- [Purushothama Rao Tata](#) ORCID: [orcid.org/0000-0003-4837-0337<sup>1,12,13,14,15</sup>](https://orcid.org/0000-0003-4837-0337)

*Nature* volume **604**, pages 111–119 (2022)

- 7007 Accesses
- 79 Altmetric

- [Metrics details](#)

## Subjects

- [Differentiation](#)
- [Regeneration](#)

## Abstract

Mapping the spatial distribution and molecular identity of constituent cells is essential for understanding tissue dynamics in health and disease. We lack a comprehensive map of human distal airways, including the terminal and respiratory bronchioles (TRBs), which are implicated in respiratory diseases<sup>1,2,3,4</sup>. Here, using spatial transcriptomics and single-cell profiling of microdissected distal airways, we identify molecularly distinct TRB cell types that have not—to our knowledge—been previously characterized. These include airway-associated *LGR5*<sup>+</sup> fibroblasts and TRB-specific alveolar type-0 (AT0) cells and TRB secretory cells (TRB-SCs). Connectome maps and organoid-based co-cultures reveal that *LGR5*<sup>+</sup> fibroblasts form a signalling hub in the airway niche. AT0 cells and TRB-SCs are conserved in primates and emerge dynamically during human lung development. Using a non-human primate model of lung injury, together with human organoids and tissue specimens, we show that alveolar type-2 cells in regenerating lungs transiently acquire an AT0 state from which they can differentiate into either alveolar type-1 cells or TRB-SCs. This differentiation programme is distinct from that identified in the mouse lung<sup>5,6,7</sup>. Our study also reveals mechanisms that drive the differentiation of the bipotent AT0 cell state into normal or pathological states. In sum, our findings revise human lung cell maps and lineage trajectories, and implicate an epithelial transitional state in primate lung regeneration and disease.

This is a preview of subscription content

## Access options

Subscribe to Nature+

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

\$29.99

monthly

[Subscribe](#)

Subscribe to Journal

Get full journal access for 1 year

\$199.00

only \$3.90 per issue

[Subscribe](#)

All prices are NET prices.

VAT will be added later in the checkout.

Tax calculation will be finalised during checkout.

Buy article

Get time limited or full article access on ReadCube.

\$32.00

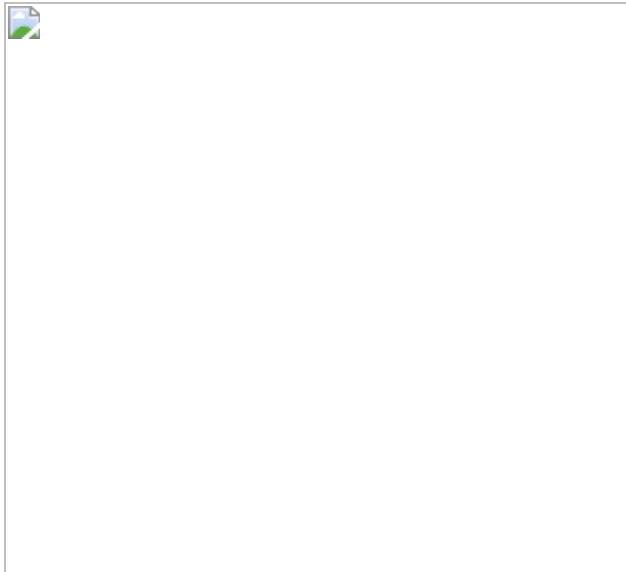
[Buy](#)

All prices are NET prices.

### **Additional access options:**

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

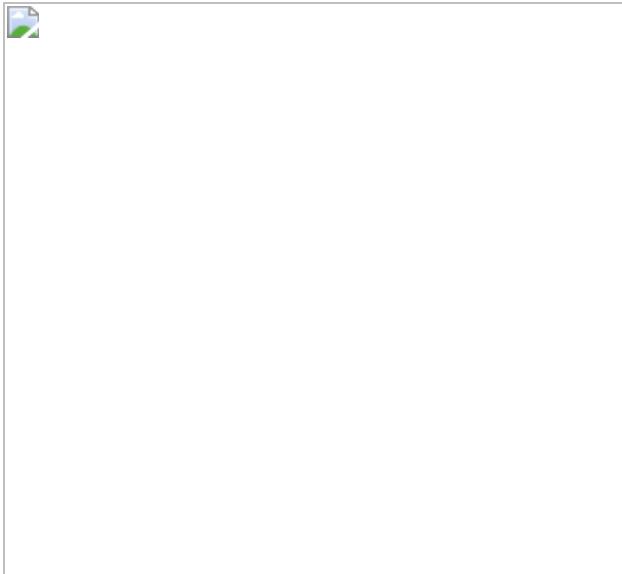
**Fig. 1: Spatial and single-cell transcriptomics reveal distinct epithelial cell types in distal airways.**



**Fig. 2: *LGR5* marks a unique fibroblast population in airways.**



**Fig. 3: Dynamics of AT0 and unique cell trajectories in human alveolar epithelium.**



**Fig. 4: Revised lineage hierarchies in primate alveolar regeneration and disease.**



## Data availability

scRNA-seq data generated in this study are available from the Gene Expression Omnibus (GEO) under accession code [GSE178360](#) (human airways) and [GSE178519](#) (organoids). Spatial transcriptomics data generated in this study are available under accession code [GSE178361](#). snRNA-seq data from monkey lungs<sup>45</sup> are available at the GSA (accession

code [CRA002577](#)). Human lung scRNA-seq datasets from Travaglini et al.<sup>11</sup> are available through Synapse (<https://www.synapse.org/#/Synapse:syn21041850>); Habermann et al.<sup>12</sup> from the GEO (accession code [GSE135893](#)); Deprez et al.<sup>23</sup> through the UCSC Cell Browser (and also in the European Genome-Phenome Archive; <https://www.ebi.ac.uk/ega/datasets/EGAD00001005714>); and Miller et al.<sup>28</sup> on the EMBL-EBI Array Express database (accession code [E-MTAB-8221](#)). [Source data](#) are provided with this paper.

## Code availability

All analyses performed in this study used pre-existing packages mentioned in the [Methods](#). No custom code was generated.

## References

1. Verleden, S. E. et al. Small airways pathology in idiopathic pulmonary fibrosis: a retrospective cohort study. *Lancet Respir. Med.* **8**, 573–584 (2020).
2. Hogg, J. C., Macklem, P. T. & Thurlbeck, W. M. The resistance of small airways in normal and diseased human lungs. *Aspen Emphysema Conf.* <https://europepmc.org/article/med/5610792> **10**, 433–441 (1967).
3. Tanabe, N. et al. Micro-computed tomography comparison of preterminal bronchioles in centrilobular and panlobular emphysema. *Am. J. Respir. Crit. Care Med.* **195**, 630–638 (2017).
4. Weibel, E. R. & Gomez, D. M. Architecture of the human lung. Use of quantitative methods establishes fundamental relations between size and number of lung structures. *Science* **137**, 577–585 (1962).
5. Kobayashi, Y. et al. Persistence of a regeneration-associated, transitional alveolar epithelial cell state in pulmonary fibrosis. *Nat. Cell Biol.* **22**, 934–946 (2020).

6. Strunz, M. et al. Alveolar regeneration through a Krt8<sup>+</sup> transitional stem cell state that persists in human lung fibrosis. *Nat. Commun.* **11**, 3559 (2020).
7. Choi, J. et al. Inflammatory signals induce AT2 cell-derived damage-associated transient progenitors that mediate alveolar regeneration. *Cell Stem Cell* **27**, 366–382 (2020).
8. Regev, A. et al. The Human Cell Atlas. *eLife* **6**, e27041 (2017).
9. Macosko, E. Z. et al. Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. *Cell* **161**, 1202–1214 (2015).
10. Klein, A. M. et al. Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells. *Cell* **161**, 1187–1201 (2015).
11. Travaglini, K. J. et al. A molecular cell atlas of the human lung from single-cell RNA sequencing. *Nature* **587**, 619–625 (2020).
12. Habermann, A. C. et al. Single-cell RNA sequencing reveals profibrotic roles of distinct epithelial and mesenchymal lineages in pulmonary fibrosis. *Sci. Adv.* **6**, eaba1972 (2020).
13. Adams, T. S. et al. Single-cell RNA-seq reveals ectopic and aberrant lung-resident cell populations in idiopathic pulmonary fibrosis. *Sci. Adv.* **6**, eaba1983 (2020).
14. Carraro, G. et al. Single-cell reconstruction of human basal cell diversity in normal and idiopathic pulmonary fibrosis lungs. *Am. J. Respir. Crit. Care Med.* **202**, 1540–1550 (2020).
15. Okuda, K. et al. Secretory cells dominate airway CFTR expression and function in human airway superficial epithelia. *Am. J. Respir. Crit. Care Med.* **203**, 1275–1289 (2021).
16. Plasschaert, L. W. et al. A single-cell atlas of the airway epithelium reveals the CFTR-rich pulmonary ionocyte. *Nature* **560**, 377–381

(2018).

17. Montoro, D. T. et al. A revised airway epithelial hierarchy includes CFTR-expressing ionocytes. *Nature* **560**, 319–324 (2018).
18. Vieira Braga, F. A. et al. A cellular census of human lungs identifies novel cell states in health and in asthma. *Nat. Med.* **25**, 1153–1163 (2019).
19. Hogg, J. C., Macklem, P. T. & Thurlbeck, W. M. Site and nature of airway obstruction in chronic obstructive lung disease. *N. Engl. J. Med.* **278**, 1355–1360 (1968).
20. Hogg, J. C. et al. The nature of small-airway obstruction in chronic obstructive pulmonary disease. *N. Engl. J. Med.* **350**, 2645–2653 (2004).
21. Cosio, M. et al. The relations between structural changes in small airways and pulmonary-function tests. *N. Engl. J. Med.* **298**, 1277–1281 (1978).
22. Weibel, E. R. A retrospective of lung morphometry: from 1963 to present. *Am. J. Physiol. Lung Cell. Mol. Physiol.* **305**, L405–L408 (2013).
23. Deprez, M. et al. A single-cell atlas of the human healthy airways. *Am. J. Respir. Crit. Care Med.* **202**, 1636–1645 (2020).
24. Buechler, M. B. et al. Cross-tissue organization of the fibroblast lineage. *Nature* **593**, 575–579 (2021).
25. Schittny, J. C. Development of the lung. *Cell Tissue Res.* **367**, 427–444 (2017).
26. Pan, H., Deutsch, G. H. & Wert, S. E. Comprehensive anatomic ontologies for lung development: a comparison of alveolar formation and maturation within mouse and human lung. *J. Biomed. Semant.* **10**, 18 (2019).

27. Jeffrey, P. K. The development of large and small airways. *Am. J. Respir. Crit. Care Med.* **157**, S174–S180 (1998).
28. Miller, A. J. et al. In vitro and in vivo development of the human airway at single-cell resolution. *Dev. Cell* **53**, 117–128 (2020).
29. La Manno, G. et al. RNA velocity of single cells. *Nature* **560**, 494–498 (2018).
30. Street, K. et al. Slingshot: cell lineage and pseudotime inference for single-cell transcriptomics. *BMC Genomics* **19**, 477 (2018).
31. Bergen, V., Lange, M., Peidli, S., Wolf, F. A. & Theis, F. J. Generalizing RNA velocity to transient cell states through dynamical modeling. *Nat. Biotechnol.* **38**, 1408–1414 (2020).
32. Cao, J. et al. The single-cell transcriptional landscape of mammalian organogenesis. *Nature* **566**, 496–502 (2019).
33. Wolf, F. A. et al. PAGA: graph abstraction reconciles clustering with trajectory inference through a topology preserving map of single cells. *Genome Biol.* **20**, 59 (2019).
34. Jin, S., MacLean, A. L., Peng, T. & Nie, Q. scEpath: energy landscape-based inference of transition probabilities and cellular trajectories from single-cell transcriptomic data. *Bioinformatics* **34**, 2077–2086 (2018).
35. Zhou, P., Wang, S., Li, T. & Nie, Q. Dissecting transition cells from single-cell transcriptome data through multiscale stochastic dynamics. *Nat. Commun.* **12**, 5609 (2021).
36. Katsura, H. et al. Human lung stem cell-based alveolospheres provide insights into SARS-CoV-2-mediated interferon responses and pneumocyte dysfunction. *Cell Stem Cell* **27**, 890–904 (2020).
37. Plantier, L. et al. Ectopic respiratory epithelial cell differentiation in bronchiolised distal airspaces in idiopathic pulmonary fibrosis. *Thorax* **66**, 651–657 (2011).

38. Weibel, E. R. *Morphometry of the Human Lung* (Springer-Verlag, 1963).
39. Castleman, W. L., Dungworth, D. L. & Tyler, W. S. Intrapulmonary airway morphology in three species of monkeys: a correlated scanning and transmission electron microscopic study. *Am. J. Anat.* **142**, 107–121 (1975).
40. Pinkerton, K. E. & Joad, J. P. The mammalian respiratory system and critical windows of exposure for children's health. *Environ. Health Perspect.* **108**, 457–462 (2000). Suppl 3.
41. Hyde, D. M., Samuelson, D. A., Blakeney, W. H. & Kosch, P. C. A correlative light microscopy, transmission and scanning electron microscopy study of the ferret lung. *Scan. Electron Microsc.* **3**, 891–898 (1979).
42. Miller, L. A., Royer, C. M., Pinkerton, K. E. & Schelegle, E. S. Nonhuman primate models of respiratory disease: past, present, and future. *ILAR J.* **58**, 269–280 (2017).
43. Rogers, C. S. et al. The porcine lung as a potential model for cystic fibrosis. *Am. J. Physiol. Lung Cell Mol. Physiol.* **295**, L240–L263 (2008).
44. Plopper, C. G., Heidsiek, J. G., Weir, A. J., George, J. A. & Hyde, D. M. Tracheobronchial epithelium in the adult rhesus monkey: a quantitative histochemical and ultrastructural study. *Am. J. Anat.* **184**, 31–40 (1989).
45. Ma, S. et al. Single-cell transcriptomic atlas of primate cardiopulmonary aging. *Cell Res.* **31**, 415–432 (2021).
46. Boucher, R. C. Muco-obstructive lung diseases. *N. Engl. J. Med.* **380**, 1941–1953 (2019).
47. Bhandari, A. & McGrath-Morrow, S. Long-term pulmonary outcomes of patients with bronchopulmonary dysplasia. *Semin. Perinatol.* **37**,

132–137 (2013).

48. Jiang, Y. et al. Alteration of cystic airway mesenchyme in congenital pulmonary airway malformation. *Sci. Rep.* **9**, 5296 (2019).
49. Kim, C. F. B. et al. Identification of bronchioalveolar stem cells in normal lung and lung cancer. *Cell* **121**, 823–835 (2005).
50. Giangreco, A., Reynolds, S. D. & Stripp, B. R. Terminal bronchioles harbor a unique airway stem cell population that localizes to the bronchoalveolar duct junction. *Am. J. Pathol.* **161**, 173–182 (2002).
51. Williams, S. E., Beronja, S., Pasolli, H. A. & Fuchs, E. Asymmetric cell divisions promote Notch-dependent epidermal differentiation. *Nature* **470**, 353–358 (2011).
52. Kumar, P. A. et al. Distal airway stem cells yield alveoli in vitro and during lung regeneration following H1N1 influenza infection. *Cell* **147**, 525–538 (2011).
53. Nagendran, M., Riordan, D. P., Harbury, P. B. & Desai, T. J. Automated cell-type classification in intact tissues by single-cell molecular profiling. *eLife* **7**, e30510 (2018).
54. Kwapiszewska, G. et al. Transcriptome profiling reveals the complexity of pirfenidone effects in idiopathic pulmonary fibrosis. *Eur. Respir. J.* **52**, 800564 (2018).
55. Stuart, T. et al. Comprehensive integration of single-cell data. *Cell* **177**, 1888–1902 (2019).
56. Young, M. D. & Behjati, S. SoupX removes ambient RNA contamination from droplet-based single-cell RNA sequencing data. *Gigascience* **9**, giaa151 (2020).
57. Hafemeister, C. & Satija, R. Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression. *Genome Biol.* **20**, 296 (2019).

58. Chen, E. Y. et al. Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinformatics* **14**, 128 (2013).
59. Lambert, S. A. et al. The human transcription factors. *Cell* **172**, 650–665 (2018).
60. Aibar, S. et al. SCENIC: single-cell regulatory network inference and clustering. *Nat. Methods* **14**, 1083–1086 (2017).
61. Jin, S. et al. Inference and analysis of cell–cell communication using CellChat. *Nat. Commun.* **12**, 1088 (2021).

## Acknowledgements

We thank B. Hogan for advice and critical reading of the manuscript; members of the P.R.T. laboratory for discussions; the Human Cell Atlas community for discussions and feedback; and members of the S.H.R. laboratory at UNC Chapel Hill for providing human tissues (supported by grant BOUCHER19R0 from the CF Foundation and NIH grant DK065988). We acknowledge support from the Duke University Molecular Genomics Core Facility (for help with Visium experiments); and from the Duke BRPC shared resource (for providing lung tissue specimens). V.S. is supported by a fellowship from Regeneration Next initiative at Duke University. Y.K. was supported by a fellowship from Japan Society for the Promotion of Science Overseas Research. Human tissue procurement was supported by funding from the CF Foundation (BOUCHER19R0) and the NIH (DK065988). J.R.S. is supported by grant CZF2019-002440 from the Chan Zuckerberg Initiative DAF, an advised fund of the Silicon Valley Community Foundation, and by the National Heart, Lung and Blood Institute (NHLBI; R01HL119215). We also thank I. Glass and the University of Washington Laboratory of Developmental Biology, which is supported by NIH award 5R24HD000836 from the Eunice Kennedy Shriver National Institute of Child Health and Human Development (NICHD). This work was supported by funds from the NIH/NHLBI (R01 HL127002, U01HL134766) to J.R.R.; the NIH (P51OD011107) to L.A.M.; the NIH (R01 DK047967-30) to J.F.E.; a Pathways to Independence award from the NHLBI/NIH (R00HL127181, R01HL146557 and R01HL153375); a

NCI/NIH P30 Cancer Center Support Grant (P30CA014236); funds from the Chan Zuckerberg Initiative–Human Lung Cell Atlas pilot award to P.R.T.; and funds from Regeneration NeXT and Kaganov–MEDx Pulmonary Initiative to P.R.T. at Duke University. This work was partially supported by funds from the Whitehead foundation and P.R.T. is a Whitehead Scholar at Duke University.

## Author information

### Author notes

1. Yoshihiko Kobayashi

Present address: Division of Pulmonary Critical Care, Department of Medicine, Duke University School of Medicine, Durham, NC, USA

2. These authors contributed equally: Preetish Kadur Lakshminarasimha Murthy, Vishwaraj Sontake

## Affiliations

1. Department of Cell Biology, Duke University School of Medicine, Durham, NC, USA

Preetish Kadur Lakshminarasimha Murthy, Vishwaraj Sontake, Aleksandra Tata, Yoshihiko Kobayashi, Lauren Macadlo & Purushothama Rao Tata

2. Marsico Lung Institute, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA

Kenichi Okuda, Satoko Nakano & Richard C. Boucher

3. Graduate Program in Cellular and Molecular Biology, University of Michigan Medical School, Ann Arbor, MI, USA

Ansley S. Conchola

4. Department of Neurology, Duke University Medical Center, Durham, NC, USA

Simon Gregory

5. California National Primate Research Center, Davis, CA, USA

Lisa A. Miller

6. Department of Anatomy, Physiology and Cell biology, School of Veterinary Medicine, University of California, Davis, CA, USA

Lisa A. Miller

7. Department of Cell and Developmental Biology, University of Michigan Medical School, Ann Arbor, MI, USA

Jason R. Spence

8. Department of Internal Medicine, Carver College of Medicine, University of Iowa, Iowa City, IA, USA

John F. Engelhardt

9. Department of Anatomy & Cell Biology, Carver College of Medicine, University of Iowa, Iowa City, IA, USA

John F. Engelhardt

10. Department of Immunology Discovery, Genentech, South San Francisco, CA, USA

Jason R. Rock

11. Department of Cell Biology and Physiology, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA

Scott H. Randell

12. Duke Cancer Institute, Duke University School of Medicine, Durham, NC, USA

Purushothama Rao Tata

13. Duke Regeneration Center, Duke University School of Medicine, Durham, NC, USA

Purushothama Rao Tata

14. Division of Pulmonary Critical Care, Department of Medicine, Duke University School of Medicine, Durham, NC, USA

Purushothama Rao Tata

15. Center for Advanced Genomic Technologies, Duke University, Durham, NC, USA

Purushothama Rao Tata

16. Institute for Frontier Life and Medical Sciences, Kyoto University, Kyoto, Japan

Yoshihiko Kobayashi

## Contributions

P.K.L.M. co-designed and performed scRNA-seq, Visium, RNAScope, qRT-PCR, plasmid cloning, FACS and non-human primate, human fetal, pediatric and adult lung characterization experiments, analysed data, performed computational analysis and co-wrote the manuscript; V.S. co-designed and performed airway dissociation, plasmid cloning, FACS, qRT-PCR, organoid culture and basal and fibroblast co-culture experiments, performed immunostaining on lung tissues from patients with acute lung injury, COPD and IPF and co-wrote the manuscript. A.T. performed microdissection and dissociation of airways, performed immunostaining on human (healthy and IPF) and non-human primate lung tissues and co-wrote the manuscript. Y.K. performed scRNA-seq and RNA-FISH (PLISH)

experiments and assisted with computational analysis. L.M. assisted with human tissue processing, sectioning of human and monkey lungs and immunostaining. K.O. performed RNAScope experiments. K.O. and S.N. performed cilia length measurements. A.S.C. and J.R.S. provided fetal lung samples. S.G. supervised tissue optimization for Visium experiments. L.A.M. and J.R.R. supervised bleomycin-induced injury in non-human primates and provided samples. J.F.E. provided insights into distal airway anatomy across species. R.C.B. supervised RNAScope experiments. S.H.R. provided human lung tissues. P.R.T. co-designed, conceived and supervised the work and co-wrote the manuscript. All authors reviewed and edited the manuscript.

## **Corresponding author**

Correspondence to [Purushothama Rao Tata](#).

## **Ethics declarations**

## **Competing interests**

P.R.T. serves as a consultant for Surrozen, Cellarity and Celldom on work not related to the contents of this manuscript. J.R.R. is an employee of Genentech. The other authors declare no competing interests.

## **Peer review**

## **Peer review information**

*Nature* thanks Qing Nie and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

## **Additional information**

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## Extended data figures and tables

### Extended Data Fig. 1 Spatial transcriptomics of cartilaginous airway and distal lung.

a. Heat map shows differentially expressed genes in each cluster of ST assay from proximal airway. b. ST spots show expression of indicated genes from proximal airway. c. Heat map indicates differentially expressed genes in each cluster of ST assay from the distal lung. d. ST spots indicate the expression of indicated genes from the distal lung.

### Extended Data Fig. 2 scRNA-seq of human distal airways and alveoli.

a. UMAPs show the expression of cell class specific marker genes. b. Dot plot indicates expression of cell type specific markers. c. UMAP shows the annotation of cell types from blood (c), endothelial (d), and mesenchymal (e) lineages. f. H&E reference sections and ST spots show the expression of *RNASE1* in proximal (top) and distal airway (bottom). g. ST spots show the expression of *MUC5AC*, *SFTPB*, *MUC5B* and *SCGB3A2* in proximal and distal airways as indicated. h. PLISH for *RNASE1* (grey), *MUC5B* (green) and *SFTPB* (red) in proximal (top) and distal airways (bottom). blue (DAPI) indicates nuclei. Scale bar: 50  $\mu$ m. i. Visualization of indicated epithelial cell transcriptomic signatures on ST tissue section from distal lung.

### Extended Data Fig. 3 Molecularly distinct cell types and zones in human airways.

a. Immunostaining for MUC5B (blue), MUC5AC (red), SCGB1A1 (green) and SCGB3A2 (grey) (bottom panel). Lines illustrate described zones. Scale bar: 100  $\mu$ m. b. Quantification of distal airway epithelial cell types across the airway zones. Data are presented as mean  $\pm$  SEM. (n = 4 for

zone-1 and n = 3 for zones-2, 3 and 4). c. Co-staining for SCGB3A2 (green), SCGB1A1 (red), SFTPB (grey) in different regions of the human airway. Scale bar: 100  $\mu$ m. d. Staining for SCGB1A1 (blue), MUC5B (green), SFTPB (red) and SCGB3A2 (grey) in different regions of the airway. Scale bar: 100  $\mu$ m (main panel) and 20  $\mu$ m (inset). e. Box plot showing the difference in cell populations identified in this study (red bar) and Habermann et al., (turquoise bar). Line in the box: median; box edges: 25/75 percentile; whiskers: 1.5 times interquartile range; points: outliers. f. Dot plot shows expression of indicated genes in TRB-epithelial cells (this study) but not in Deprez et al., scRNA-seq data from airway epithelial cells (annotated as distal airway). In panels c and d, images were stitched to show large area. Insets show zoomed images of regions highlighted in boxes. g. UMAP plots show the expression of indicated genes in epithelial cells. h. UMAP plots show the expression of indicated genes in epithelial cells. i. (top) Immunostaining for SCGB3A2 (green), SCGB1A1 (grey) and NAPSA (red) showing the expression of NAPSA in TRB-SC and TB-SC. Scale bar: 10 and 20  $\mu$ m. (bottom) RNA-FISH for *SCGB3A2* (green), *TMEM45A* (red) and *SCGB1A1* (grey) showing the expression of *TMEM45A* in TRB-SC and TB-SC. blue (DAPI) indicates nuclei. Scale bar: 20  $\mu$ m. j. PLISH for *SCGB3A2* (green), *SFTPB* (red) and *KDR* (grey). Scale bar 50  $\mu$ m. k. UMAPs show the expression of indicated genes in epithelial cell types. l. Bar plot showing the pathways enriched in TRB-SCs.

## [Source Data](#)

## [\*\*Extended Data Fig. 4 Gene expression signatures and pathway analysis of AT0s, distal-BCs and SCGB3A2-CCs.\*\*](#)

a. Volcano plot shows differential gene expression between TRB-SC and AT0s. b. Volcano plot shows differential gene expression between AT2 and AT0s. c. Immunostaining for ACTA2 (green), SFTPB (red) and SCGB3A2 (grey) (left and top right) showing rare AT0s found in alveolar septae in healthy lungs. Immunostaining for SFTPB (green), SFTPC (red) and SCGB3A2 (grey) (bottom right) showing alveolar sacs. blue (DAPI) indicates nuclei. Scale bar: 50 and 20  $\mu$ m. d. UMAP plot and immunostaining for SOX2 (green), SFTPC (red) show the expression of SOX2 in AT0s near respiratory bronchioles. Scale bar: 50  $\mu$ m. Inset:

individual channels of regions indicated by a black box. e. Violin plots show expression patterns of genes encoding indicated transcription factors in TRB-SC, AT0, and AT2s. f. Heat map shows expression of indicated genes in deuterosomal, ciliated, SCGB3A2-CC and TRB-SC cell populations. (Scale shows z-score). g. Volcano plot shows differential gene expression between SCGB3A2-CC and deuterosomal cell populations. h. Quantification of SCGB3A2-CC cell type across the airway zones. Data are presented as mean  $\pm$  SEM ( $n = 4$  for zones-1, 2 and 4;  $n = 3$  for zone-3). i. Heat map shows expression of indicated genes in basal, distal-BC-1 and distal-BC-2 cell populations. (rows were hierarchically clustered and scale shows z-score). j. Violin plot shows the expression of indicated genes in basal cell populations. k. Violin plot demonstrates the expression of indicated genes in different basal cell populations. l. Quantification of distal basal cell types across the airway zones. Data are presented as mean  $\pm$  SEM ( $n = 4$  for zone-1,  $n = 5$  for zone-2,  $n = 3$  for zone-3 and  $n = 4$  for zone-4). m. Volcano plot shows differential gene expression between IPF basaloid (from Habermann et al.,) and distal-BC-2 cell populations. n and o. UMAPs show the expression of *TP63*, *SFTPB* and *FOXJ1* and *SCGB3A2* in distal, intermediate and proximal airway cells profiled in Deprez et al.

#### [Source Data](#)

#### **Extended Data Fig. 5 Transcriptional regulation of TRB-specific epithelial cells.**

- a. Heat map showing the enrichment of transcription factor regulons in epithelial cell types. b. Dot plots indicate the expression of transcription factors in epithelial cell types.

#### **Extended Data Fig. 6 Molecular and phenotypic differences of basal cells isolated from proximal and distal airways.**

- a. Schematic depicting the location of airway for sampling the basal cells (top, proximal airways and bottom, distal airways) for organoid cultures and ALI cultures. b. Representative basal cell organoid images of proximal and distal airways. c. Quantification of the colony formation efficiency (CFE). Data are presented as mean  $\pm$  SEM. NS ( $p = 0.5462$ ) (unpaired two-tailed t-

test) ( $n = 3$ ). d. SFTPB (green), TP63 (red) and CK5 expression (grey) on 2D cultures. blue (DAPI) indicates nuclei. Scale bar 50  $\mu\text{m}$ . e. Representative whole mount immunostaining image for SCGB3A2 (green), SCGB1A1 (red) and SFTPB (red) expression in basal cell ALI culture of proximal and distal airways at day 28. blue (DAPI) indicates nuclei. Scale bar: 20  $\mu\text{m}$ . f. Relative mRNA expression of *TP63*, *SFTPB*, *KRT5* and *RNASE1* in basal cell cultures. Data are presented as mean  $\pm$  SEM. (\*\*  $p = 0.0073$  for *TP63*), (\*  $p = 0.0490$  for *KRT5*), (\*  $p = 0.0130$  for *SFTPB*) and (\*\*  $p = 0.0006$  for *RNASE1*) (unpaired two-tailed t-test) ( $n = 3$ , distal airway;  $n = 4$ , proximal airway, *SFTPB* mRNA was not detected in two biological replicates of proximal airway). g. Relative mRNA expression of *TP63*, *SFTPB*, *RNASE1*, *SCGB1A1*, *SCGB3A2*, and *FOXJ1* in basal cell ALI culture of proximal and distal airways at day 28. Data are presented as mean  $\pm$  SEM. (\*  $p = 0.0312$  for *TP63*), (\*\*  $p = 0.0003$  for *SFTPB*), (\*\*  $p = 0.0028$  for *RNASE1*), (\*  $p = 0.0234$  for *SCGB3A2*), (\*\*  $p = 0.0071$  for *SCGB1A1*), and (NS  $p = 0.75$  for *FOXJ1*) (unpaired two-tailed t-test with Welch's correction) ( $n = 5$ , distal airway;  $n = 4$ , proximal airway). h. Representative histology images showing the difference in ciliary length in proximal and distal airways. Scale bar: 10  $\mu\text{m}$ . i. Representative histology images showing the difference in ciliary length in basal cell ALI culture proximal and distal airways at day 28. Scale bar: 10  $\mu\text{m}$ . j. Quantification of cilia length and cilia beating frequency in basal cell ALI culture of proximal and distal airways. Data are presented as mean  $\pm$  SEM. \* ( $p = 0.0359$ ) (unpaired two-tailed t-test) ( $n = 6$ ).

### Source Data

### Extended Data Fig. 7 Transcriptional and signalling activity of newly identified mesenchymal cells.

a. ST spots show the expression of *FGF14* and *WNT2*. b. Representative RNA-FISH images showing the expression of *SFTPB* (green), *LGR5* (red) and *PDGFRA* (grey) (left) and *SCGB3A2* (green), *LGR5* (red) and *PDGFRA* (grey) (right) and Scale bar: 50  $\mu\text{m}$  and 20  $\mu\text{m}$ . DAPI stains nuclei (blue). Insets show magnified regions. c. Quantification of *LGR5*<sup>+</sup> fibroblasts in large and distal airway. Data are presented as mean  $\pm$  SEM. NS ( $p = 0.3152$ ) (unpaired two-tailed t-test) ( $n = 3$ ). d. Heat map showing net

outgoing and incoming signals in respective cell types as predicted by the CellChat algorithm. e. FGF and PDGF signalling activity from (top panel) and towards (bottom panel) *LGR5*<sup>+</sup> cells and their interaction with epithelial cells as predicted using the CellChat algorithm. f. Heat map shows the enrichment of transcription factor regulons in mesenchymal cell types. g. Heat map shows *LGR5*<sup>+</sup> fibroblast specific transcription factor expression pattern in mesenchymal cell types (scale shows z-score).

[Source Data](#)

**Extended Data Fig. 8 *LGR5*-mRFP reporter validation and co-culture.**

a. Schematic depicting pLKO-*LGR5*-mRFP reporter construct. b. Representative gating for FACS sorting of *LGR5*-mRFP reporter fibroblasts used for organoid co-cultures. c. RNA-FISH for *LGR5* (left), and *PDGFRα* (right) on *LGR5*-mRFP fibroblasts. d. Relative mRNA expression of *LGR5*, *WNT5A*, *WNT2*, *CA3*, and *FGF7* in cultured *LGR5*-mRFP fibroblasts vs alveolar fibroblasts. Data are presented as mean ± SEM. (\*\*  $p = 0.0055$  for *LGR5*), (\*  $p = 0.0187$  for *WNT5A*), (NS  $p = 0.0972$  for *CA3*), (NS  $p = 0.1216$  for *WNT2*), and (NS  $p = 0.1525$  for *FGF7*) (unpaired two-tailed t-test with Welch's correction) ( $n = 3$ ). e. Immunostaining for SFTPB (green), TP63 (red) and KRT5 (grey) in basal cell organoid co-cultured with *LGR5*-mRFP fibroblasts or alveolar fibroblasts. blue (DAPI) indicates nuclei. Scale bar: 50 μm. f. Relative mRNA expression of *SCGB3A2* and *SOX2* in AT2s co-cultured with *LGR5*-mRFP fibroblasts. Data are presented as mean ± SEM. NS  $p = 0.2383$ , and \* $p = 0.0383$  and (unpaired two-tailed t-test with Welch's correction) ( $n = 3$ ).

**Extended Data Fig. 9 Expression pattern of markers of distal airway cells in fetal and postnatal human airways.**

a. UMAP shows expression pattern of indicated genes in fetal lung epithelial scRNA-seq data from Miller et al. b. Schematic representation of human fetal airway and distal tip bud. c. Immunostaining for *SCGB3A2* (green) and SFTPB (grey) on 19-20 weeks old human fetal lung. Scale bar

100 μm. Insets show magnified regions. d. Immunostaining for SCGB3A2 (green), TP63 (red) and SFTPB (grey) in human fetal extrapulmonary airway. Scale bar: 100 μm. e. Representative RNA-FISH images showing the expression of *SCGB3A2* (green), *LGR5* (red) and *PDGFRA* (grey). Scale bar: 50 μm. DAPI stains nuclei (blue). f. Schematic representation of human postnatal lung showing TRBs and distal alveoli. g. Co-staining for SCGB3A2 (green) and SFTPB (grey) in 7-month-old postnatal human lung. Scale bar 100 μm. Insets show magnified regions. h. Immunostaining for SCGB1A1 (blue), SCGB3A2 (grey) and SFTPB (green) expression in postnatal human lung. Scale bar: 100 μm. i. Co-staining for SCGB3A2 (green), TP63 (red) and SFTPB (grey) expression in 7–12 months old postnatal lung. Scale bar: 100 μm. j–m. Representative immunostaining images from 7 months old postnatal human lung show the expression of (j) KRT5 (green), TP63 (red) and SFTPB (grey). Scale bars: 20 μm (top), 10 μm (bottom left) and 5 μm (bottom right). (k) SFTPB (green), FOXJ1 (red) and SCGB3A2 (grey). Scale bar: 10 μm. (l) SCGB3A2 (green) and SFTPB (red). Scale bar 10 μm. (m) SCGB3A2 (green) and SFTPC (red). Scale bar 20 μm. DAPI stains nuclei (blue). c, e–m: Insets indicate individual colour channels of regions marked in white boxes.

## **Extended Data Fig. 10 Computational prediction of AT2 cellular transition dynamics.**

a. PAGA graph of AT0, AT1, Immature AT1, AT2 and TRB-SC cells from healthy adult human lung. b. Cellular trajectory on AT0, AT1, Immature AT1, AT2 and TRB-SC cells as inferred by scVelo and PAGA algorithms. c. Cellular trajectory (left) and cells ordered along pseudotime (right) as inferred by Monocle 3 algorithm on of AT0, AT1, Immature AT1, AT2 and TRB-SC cells. d. Overall energy landscape view of AT0, AT1, Immature AT1, AT2 and TRB-SC cells as predicted by scEpath algorithm. e. MuTrans analysis of AT0, AT1, Immature AT1, AT2 and TRB-SC cells. Plots showing (from left to right) entropy, SFTPC level and SCGB3A2 level. f. Heat map showing expression of genes along trajectory from AT2 to AT1 predicted by Slingshot algorithm. Genes enriched in AT0s are highlighted by the black outline. g. Line plots showing smoothed expression (Loess smoothed) of select genes along Slingshot predicted AT2 to AT1 trajectory. Grey indicates 95% confidence interval derived from the

smoothened fit. h. Line plots showing smoothened expression (Loess smoothened) of select genes along Slingshot predicted AT2 to TRB-SC trajectory. Grey indicates 95% confidence interval derived from the smoothened fit.

### **Extended Data Fig. 11 Organoid models reveal cell trajectories in human lungs.**

a. Schematic depicting experimental workflow for signalling modulation in organoid cultures. b. Representative immunofluorescence images show expression of SCGB3A2 (green), SFTPB (red), and SFTPC (grey) in organoids cultured in CHIR, FGF10, SB43152 and EGF depletion conditions for 21 days. Scale bar: 20  $\mu$ m. c. Staining for SOX2 (green), SFTPB (red), and SFTPC (grey) in control and EGF depleted organoids. Scale bar: 20  $\mu$ m. d. Staining for SFTPC (green), SFTPB (red), and SCGB3A2 (grey) expression in organoid cultures treated with Erlotinib (2  $\mu$ m) at day 7. Scale bar: 20  $\mu$ m. e. Relative mRNA expression of *SFTPC*, and *SCGB3A2* in AT2 cultures treated with indicated small molecules for 7 days. Data are presented as mean  $\pm$  SEM. \*\*\* ( $p = 0.0005$  for GSK690693), NS ( $p = 0.6722$  for BAY73-4506), NS ( $p = 0.2167$  for PD0325901), NS ( $p = 0.1279$  for SP600125) for *SFTPC*. \* ( $p = 0.0133$  for GSK690693), \*\* ( $p = 0.0081$  for BAY73-4506), \*\*\* ( $p = 0.0003$  for PD0325901), NS ( $p = 0.0940$  for SP600125) for *SCGB3A2*. (Dunnett's multiple comparisons test) ( $n = 3$ ). f. (left) Schematic depicting the experimental workflow to study AT2 to AT1 differentiation. (right) Immunostaining for SCGB3A2 (green) and HTI-56 (red), and SFTPB (grey) on organoids cultured in ADM for 6 days. Scale bar: 20  $\mu$ m. DAPI stains nuclei (blue). g. Representative gating for FACS sorting of AT0s based on HTI-56 levels. h. Immunostaining for SCGB3A2 (green) and HTI-56 (red), and SFTPC (grey) on AT0s cultured in ADM for 7 days.

### **Extended Data Fig. 12 scRNA-seq on human lung organoids from different treatments and AT0 sorting.**

a. Schematic showing experimental workflow for organoid scRNA-seq. b. UMAPs show expression pattern of indicated genes in specific cell

populations as revealed by scRNA-seq from different culture conditions. c. Scatterplot showing the relative expression of genes in AT0s when compared to AT2s in organoid cultures (EGF depletion or serum addition) or *in vivo*. Linear regression is shown by the black line and Pearson correlation is shown with 95% confidence interval indicated by grey colour.

### **Extended Data Fig. 13 TRB regions of non-human primates.**

a. Immunostaining for SFTPB (green) and SCGB3A2 (red) in terminal and respiratory bronchioles of Rhesus macaque. Scale bar 100  $\mu\text{m}$ . Images from adjacent regions were aligned to show large area. b. Representative immunofluorescence images from terminal and respiratory bronchioles of Rhesus macaque show the expression of KRT5 (green), TP63 (red) and SFTPB (grey). Scale bar 12.5  $\mu\text{m}$  (left and right) and 15  $\mu\text{m}$  (middle). c. Quantification of AT0s in TRB regions in control and bleomycin treated lungs. Data are presented as mean  $\pm$  SEM ( $p = 0.496$ ) (unpaired two-tailed t-test) ( $n = 3$ ). d. Immunostaining for SCGB3A2 (green), SFTPB (red), and SFTPC (grey) in COPD lung. Scale bar: 50  $\mu\text{m}$ . (left). e. Staining for SCGB3A2 (green), ACTA2 (red), and SFTPC (grey) in mild fibrotic (left) and bronchiolized region (right) of the IPF lung. Scale bar: 50  $\mu\text{m}$ . f. Staining for SOX2 (green) and SFTPC (red) in mild fibrotic (left) and bronchiolized region (right) of the IPF lung. Scale bar: 50  $\mu\text{m}$ . Maximum intensity projections of z-stacks are shown. g. Staining for SCGB3A2 (green), KI67 (red) and SFTPC (grey) in healthy human lung and in acute lung injury. Scale bar: 50  $\mu\text{m}$ . h. Quantification of proliferating ( $\text{Ki67}^+$ ) cells in control and acute lung injury lungs. Data are presented as mean  $\pm$  SEM ( $n = 3$ ). In panels – d-g, insets indicate individual colour channels of regions marked in white boxes.

### **Source Data**

## **Supplementary information**

### **Supplementary Figure 1**

FACS gating strategy. **a**, Gating strategy to sort *LGR5*-mRFP cells. **b**, Gating strategy to sort alveolar fibroblasts. Right panel shows unstained control. **c**, Gating strategy to sort AT0 cells from serum treated AT2 cultures.

## Reporting Summary

## Peer Review File

## Supplementary Table 1

Enriched markers in all cell clusters. Each tab contains a cell class. Non-parametric Wilcoxon rank-sum test was used with Bonferroni correction for multiple comparisons.

## Supplementary Table 2

Human tissue metadata. Age, gender, smoking history, existing lung disease, cause of death, tissue use and enzymes used for dissociation are provided.

## Supplementary Table 3

Probe sequences used for PLISH. Probe names and gene-specific and detection-probe-specific sequences are provided.

## Supplementary Table 4

qPCR primers and *LGR5* promoter PCR primers. Primer names and forward and reverse sequences are provided.

## Source data

## Source Data Fig. 2

[\*\*Source Data Fig. 3\*\*](#)

[\*\*Source Data Fig. 4\*\*](#)

[\*\*Source Data Extended Data Fig. 3\*\*](#)

[\*\*Source Data Extended Data Fig. 4\*\*](#)

[\*\*Source Data Extended Data Fig. 6\*\*](#)

[\*\*Source Data Extended Data Fig. 7\*\*](#)

[\*\*Source Data Extended Data Fig. 13\*\*](#)

## Rights and permissions

[\*\*Reprints and Permissions\*\*](#)

## About this article

### Cite this article

Kadur Lakshminarasimha Murthy, P., Sontake, V., Tata, A. *et al.* Human distal lung maps and lineage hierarchies reveal a bipotent progenitor. *Nature* **604**, 111–119 (2022). <https://doi.org/10.1038/s41586-022-04541-3>

- Received: 21 June 2021
- Accepted: 09 February 2022
- Published: 30 March 2022
- Issue Date: 07 April 2022
- DOI: <https://doi.org/10.1038/s41586-022-04541-3>

## Share this article

Anyone you share the following link with will be able to read this content:

[Get shareable link](#)

Sorry, a shareable link is not currently available for this article.

[Copy to clipboard](#)

Provided by the Springer Nature SharedIt content-sharing initiative

---

This article was downloaded by **calibre** from <https://www.nature.com/articles/s41586-022-04541-3>

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

- Article
- [Published: 30 March 2022](#)

# Human distal airways contain a multipotent secretory cell that can regenerate alveoli

- [Maria C. Basil<sup>1,2</sup> na1](#),
- [Fabian L. Cardenas-Diaz<sup>1,2</sup> na1](#),
- [Jaymin J. Kathiriya](#) [ORCID: orcid.org/0000-0003-1807-8570<sup>3</sup>](#),
- [Michael P. Morley<sup>1,2,5</sup>](#),
- [Justine Carl<sup>1,2</sup>](#),
- [Alexis N. Brumwell<sup>3</sup>](#),
- [Jeremy Katzen<sup>1,2</sup>](#),
- [Katherine J. Slovik<sup>1,2,5</sup>](#),
- [Apoorva Babu<sup>1,2,5</sup>](#),
- [Su Zhou<sup>1,2</sup>](#),
- [Madison M. Kremp<sup>1,2</sup>](#),
- [Katherine B. McCauley](#) [ORCID: orcid.org/0000-0002-8817-7120<sup>6</sup>](#),
- [Shanru Li<sup>1,2</sup>](#),
- [Joseph D. Planer<sup>1,2</sup>](#),
- [Shah S. Hussain<sup>7</sup>](#),
- [Xiaoming Liu](#) [ORCID: orcid.org/0000-0002-3902-9105<sup>8</sup>](#),
- [Rebecca Windmueller<sup>2,11</sup>](#),
- [Yun Ying<sup>1,2</sup>](#),
- [Kathleen M. Stewart<sup>1,2</sup>](#),
- [Michelle Oyster<sup>1</sup>](#),
- [Jason D. Christie<sup>1,2</sup>](#),
- [Joshua M. Diamond<sup>1</sup>](#),
- [John F. Engelhardt](#) [ORCID: orcid.org/0000-0003-2389-9277<sup>8</sup>](#),

- [Edward Cantu](#)<sup>2,9</sup>,
- [Steven M. Rowe](#) [ORCID: orcid.org/0000-0001-9045-0133](#)<sup>7</sup>,
- [Darrell N. Kotton](#) [ORCID: orcid.org/0000-0002-9604-8476](#)<sup>6,10</sup>,
- [Harold A. Chapman](#) [ORCID: orcid.org/0000-0002-5280-5527](#)<sup>3,4</sup> &
- [Edward E. Morrisey](#) <sup>1,2,11</sup>

[\*Nature\*](#) volume **604**, pages 120–126 (2022)

- 7935 Accesses
- 481 Altmetric
- [Metrics details](#)

## Subjects

- [Mechanisms of disease](#)
- [Regeneration](#)

## Abstract

The human lung differs substantially from its mouse counterpart, resulting in a distinct distal airway architecture affected by disease pathology in chronic obstructive pulmonary disease. In humans, the distal branches of the airway interweave with the alveolar gas-exchange niche, forming an anatomical structure known as the respiratory bronchioles. Owing to the lack of a counterpart in mouse, the cellular and molecular mechanisms that govern respiratory bronchioles in the human lung remain uncharacterized. Here we show that human respiratory bronchioles contain a unique secretory cell population that is distinct from cells in larger proximal airways. Organoid modelling reveals that these respiratory airway secretory (RAS) cells act as unidirectional progenitors for alveolar type 2 cells, which are essential for maintaining and regenerating the alveolar niche. RAS cell lineage differentiation into alveolar type 2 cells is regulated by Notch and Wnt signalling. In chronic obstructive pulmonary disease, RAS cells are altered transcriptionally, corresponding to abnormal alveolar type 2 cell states, which are associated with smoking exposure in both humans and

ferrets. These data identify a distinct progenitor in a region of the human lung that is not found in mouse that has a critical role in maintaining the gas-exchange compartment and is altered in chronic lung disease.

This is a preview of subscription content

## Access options

Subscribe to Nature+

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

\$29.99

monthly

[Subscribe](#)

Subscribe to Journal

Get full journal access for 1 year

\$199.00

only \$3.90 per issue

[Subscribe](#)

All prices are NET prices.

VAT will be added later in the checkout.

Tax calculation will be finalised during checkout.

Buy article

Get time limited or full article access on ReadCube.

\$32.00

[Buy](#)

All prices are NET prices.

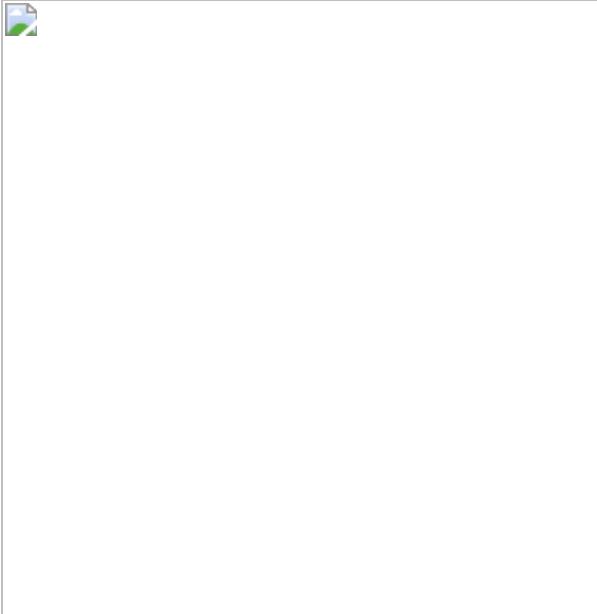
### Additional access options:

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

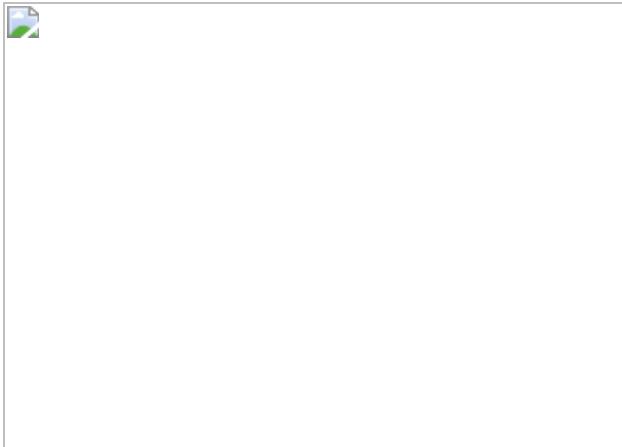
**Fig. 1: Identification of secretory cell heterogeneity in the distal human lungs.**



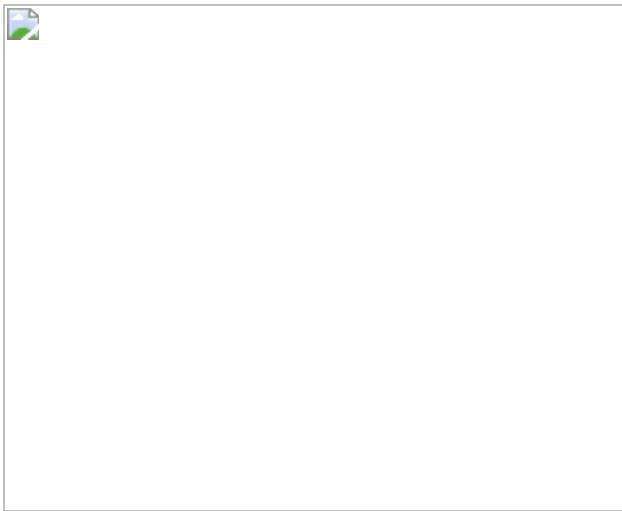
**Fig. 2: The iRAS cell transition to iAT2 cells is rapid, unidirectional, and regulated in part by Notch and Wnt signalling.**



**Fig. 3: SCGB3A2<sup>+</sup> cells, identified by CEACAM6, can be isolated from distal lung parenchyma and give rise to SFTPC<sup>+</sup> organoids ex vivo.**



**Fig. 4: COPD and exposure to cigarette smoke is associated with alternative AT2 cellular states.**



## Data availability

All scRNA-seq data generated during this study have been deposited at the Gene Expression Omnibus database (accession numbers [GSE168191](#) and [GSE130076](#)). The scRNA-seq datasets can be viewed online (<http://bit.ly/2O45FIb>). All cell lines and other reagents will be distributed on request. [Source data](#) are provided with this paper.

## Code availability

Analysis associated with the current submission used published R packages and a custom R package, which are available at GitHub

(<https://github.com/Morriseylab/scExtras>). The code for the custom graphical scRNA-seq interface is available at GitHub (<https://github.com/Morriseylab/scViewer-Lite>). For our analysis using published R packages, the following versions were used: R v.4.0.0, CellRanger v.3.1.0, Slingshot v.1.6.1, tradeSeq v.1.2.01, clusterProfiler v.3.16.1, ComplexHeatmap v.2.4.3, MAST v.1.14.0, Seurat v.3.0.1, Sctransform v.0.3.0 and clustree v.0.4.3.

## References

1. Burney, P. G., Patel, J., Newson, R., Minelli, C. & Naghavi, M. Global and regional trends in COPD mortality, 1990-2010. *Eur. Respir. J.* **45**, 1239–1247 (2015).
2. Basil, M. C. et al. The cellular and physiological basis for lung repair and regeneration: past, present, and future. *Cell Stem Cell* **26**, 482–502 (2020).
3. Zepp, J. A. & Morrisey, E. E. Cellular crosstalk in the development and regeneration of the respiratory system. *Nat. Rev. Mol. Cell Biol.* **20**, 551–566 (2019).
4. Hogan, B. L. et al. Repair and regeneration of the respiratory system: complexity, plasticity, and mechanisms of lung stem cell function. *Cell Stem Cell* **15**, 123–138 (2014).
5. Weibel, E. R., Sapoval, B. & Filoche, M. Design of peripheral airways for efficient gas exchange. *Respir. Physiol. Neurobiol.* **148**, 3–21 (2005).
6. Weibel, E. R. & Gomez, D. M. Architecture of the human lung. Use of quantitative methods establishes fundamental relations between size and number of lung structures. *Science* **137**, 577–585 (1962).
7. Have-Opbroek, A. A. W., Otto-Verberne, C. J. M., Dubbeldam, J. A. & Dijkman, J. H. The proximal border of the human respiratory unit, as

shown by scanning and transmission electron microscopy and light microscopical cytochemistry. *Anat. Rec.* **229**, 339–354 (1991).

8. Rock, J. R., Randell, S. H. & Hogan, B. L. Airway basal stem cells: a perspective on their roles in epithelial homeostasis and remodeling. *Dis. Models Mech.* **3**, 545–556 (2010).
9. Mercer, R. R., Russell, M. L., Roggeli, V. L. & Crapo, J. D. Cell number and distribution in human and rat airways. *Am. J. Respir. Cell Mol. Biol.* **10**, 613–624 (1994).
10. Ryu, J. H., Myers, J. L. & Swensen, S. J. Bronchiolar disorders. *Am. J. Respir. Crit. Care Med.* **168**, 1277–1292 (2003).
11. Hogg, J. C., Macklem, P. T. & Thurlbeck, W. M. Site and nature of airway obstruction in chronic obstructive lung disease. *New Engl. J. Med.* **278**, 1355–1360 (1968).
12. Koo, H.-K. K. et al. Small airways disease in mild and moderate chronic obstructive pulmonary disease: a cross-sectional study. *Lancet Respir. Med.* **6**, 591–602 (2018).
13. Visscher, D. W. & Myers, J. L. Bronchiolitis: the pathologist's perspective. *Proc. Am. Thorac. Soc.* **3**, 41–47 (2006).
14. Fretzayas, A. & Moustaki, M. Etiology and clinical features of viral bronchiolitis in infancy. *World J. Pediatr.* **13**, 293–299 (2017).
15. Gómez, R., Colás, C., Sebastián, A. & Arribas, J. Respiratory repercussions in adults with a history of infantile bronchiolitis. *Ann. Allergy Asthma Immunol.* **93**, 447–451 (2004).
16. Verleden, S. E. et al. Small airway loss in the physiologically ageing lung: a cross-sectional study in unused donor lungs. *Lancet Respir. Med.* **9**, 167–174 (2020).
17. Montoro, D. T. et al. A revised airway epithelial hierarchy includes CFTR-expressing ionocytes. *Nature* **560**, 319–324 (2018).

18. Plasschaert, L. W. et al. A single-cell atlas of the airway epithelium reveals the CFTR-rich pulmonary ionocyte. *Nature* **560**, 377–381 (2018).
19. Hyde, D. M., Samuelson, D. A., Blakeney, W. H. & Kosch, P. C. A correlative light microscopy, transmission and scanning electron microscopy study of the ferret lung. *Scan. Electron Microsc.* **3**, 891–898 (1979).
20. Sterner-Kock, A., Kock, M., Braun, R. & Hyde, D. M. Ozone-induced epithelial injury in the ferret is similar to nonhuman primates. *Am. J. Respir. Crit. Care Med.* **162**, 1152–1156 (2000).
21. Vinegar, A., Sinnett, E. E., Kosch, P. C. & Miller, M. L. Pulmonary physiology of the ferret and its potential as a model for inhalation toxicology. *Lab. Anim. Sci.* **35**, 246—250 (1985).
22. Bal, H. S. & Ghoshal, N. G. Morphology of the terminal bronchiolar region of common laboratory mammals. *Lab. Anim.* **22**, 76–82 (1988).
23. Finak, G. et al. MAST: a flexible statistical framework for assessing transcriptional changes and characterizing heterogeneity in single-cell RNA sequencing data. *Genome Biol.* **16**, 278 (2015).
24. Barkauskas, C. E. et al. Type 2 alveolar cells are stem cells in adult lung. *J. Clin. Invest.* **123**, 3025–3036 (2013).
25. Nabhan, A. N., Brownfield, D. G., Harbury, P. B., Krasnow, M. A. & Desai, T. J. Single-cell Wnt signaling niches maintain stemness of alveolar type 2 cells. *Science* **359**, 1118–1123 (2018).
26. Zacharias, W. J. et al. Regeneration of the lung alveolus by an evolutionarily conserved epithelial progenitor. *Nature* **555**, 251–255 (2018).
27. Street, K. et al. Slingshot: cell lineage and pseudotime inference for single-cell transcriptomics. *BMC Genom.* **19**, 477 (2018).

28. Rock, J. R. et al. Basal cells as stem cells of the mouse trachea and human airway epithelium. *Proc. Natl Acad. Sci. USA* **106**, 12771–12775 (2009).
29. McCauley, K. B. et al. Single-cell transcriptomic profiling of pluripotent stem cell-derived SCGB3A2<sup>+</sup> airway epithelium. *Stem Cell Rep.* **10**, 1579–1595 (2018).
30. McCauley, K. B. et al. Efficient derivation of functional human airway epithelium from pluripotent stem cells via temporal regulation of Wnt signaling. *Cell Stem Cell* **20**, 844–857 (2017).
31. Jacob, A. et al. Differentiation of human pluripotent stem cells into functional lung alveolar epithelial cells. *Cell Stem Cell* **21**, 472–488 (2017).
32. Miller, A. J. et al. In vitro and in vivo development of the human airway at single-cell resolution. *Dev. Cell* **53**, 117–128 (2020).
33. 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).
34. Jacob, A. et al. Derivation of self-renewing lung alveolar epithelial type II cells from human pluripotent stem cells. *Nat. Protoc.* **14**, 3303–3332 (2019).
35. Hawkins, F. et al. Prospective isolation of NKX2-1-expressing human lung progenitors derived from pluripotent stem cells. *J. Clin. Invest.* **127**, 2277–2294 (2017).
36. Guseh, J. S. et al. Notch signaling promotes airway mucous metaplasia and inhibits alveolar development. *Development* **136**, 1751–1759 (2009).
37. Morimoto, M., Nishinakamura, R., Saga, Y. & Kopan, R. Different assemblies of Notch receptors coordinate the distribution of the major

- bronchial Clara, ciliated and neuroendocrine cells. *Development* **139**, 4365–4373 (2012).
38. Rock, J. R. et al. Notch-dependent differentiation of adult airway basal stem cells. *Cell Stem Cell* **8**, 639–648 (2011).
  39. Stupnikov, M. R., Yang, Y., Mori, M., Lu, J. & Cardoso, W. V. Jagged and Delta-like ligands control distinct events during airway progenitor cell differentiation. *eLife* **8**, e50487 (2019).
  40. Tsao, P. N. et al. Notch signaling controls the balance of ciliated and secretory cell fates in developing airways. *Development* **136**, 2297–2307 (2009).
  41. Geling, A., Steiner, H., Willem, M., Bally-Cuif, L. & Haass, C. A  $\gamma$ -secretase inhibitor blocks Notch signaling in vivo and causes a severe neurogenic phenotype in zebrafish. *EMBO Rep.* **3**, 688–694 (2002).
  42. Frank, D. B. et al. Emergence of a wave of Wnt signaling that regulates lung alveogenesis by controlling epithelial self-renewal and differentiation. *Cell Rep.* **17**, 2312–2325 (2016).
  43. Zepp, J. A. et al. Distinct mesenchymal lineages and niches promote epithelial self-renewal and myofibrogenesis in the lung. *Cell* **170**, 1134–1148 (2017).
  44. Chapin, C. et al. Distribution and surfactant association of carcinoembryonic cell adhesion molecule 6 in human lung. *Am. J. Physiol. Lung Cell. Mol. Physiol.* **302**, L216–L225 (2012).
  45. Lin, S. E. et al. Expression of human carcinoembryonic antigen-related cell adhesion molecule 6 and alveolar progenitor cells in normal and injured lungs of transgenic mice. *Physiol. Rep.* **3**, e12657 (2015).
  46. Shikotra, A. et al. A CEACAM6-high airway neutrophil phenotype and ceacam6-high epithelial cells are features of severe asthma. *J. Immunol.* **198**, 3307–3317 (2017).

47. Lin, V. Y. et al. Excess mucus viscosity and airway dehydration impact COPD airway clearance. *Eur. Respir. J.* **55**, 1900419 (2020).
48. Raju, S. V. et al. A ferret model of COPD-related chronic bronchitis. *JCI Insight* **1**, e87536 (2016).
49. Stanford, D. et al. Airway remodeling in ferrets with cigarette smoke induced COPD using microCT Imaging. *Am. J. Physiol. Lung Cell. Mol. Physiol.* **319**, L11–L20 (2020).
50. Ganguly, K. et al. Secreted phosphoprotein 1 is a determinant of lung function development in mice. *Am. J. Respir. Cell Mol.* **51**, 637–651 (2014).
51. Calabrese, D. R. et al. Dectin-1 genetic deficiency predicts chronic lung allograft dysfunction and death. *JCI Insight* **4**, e133083 (2019).
52. Regeling, A. et al. HSPA6 is an ulcerative colitis susceptibility factor that is induced by cigarette smoke and protects intestinal epithelial cells by stabilizing anti-apoptotic Bcl-XL. *Biochim. Biophys. Acta* **1862**, 788–796 (2016).
53. Kim, C. F. et al. Identification of bronchioalveolar stem cells in normal lung and lung cancer. *Cell* **121**, 823–835 (2005).
54. Choi, J. et al. Release of Notch activity coordinated by IL-1 $\beta$  signalling confers differentiation plasticity of airway progenitors via Fosl2 during alveolar regeneration. *Nat. Cell Biol.* **23**, 953–966 (2021).
55. Todd, J. L. & Palmer, S. M. Bronchiolitis obliterans syndrome: the final frontier for lung transplantation. *Chest* **140**, 502–508 (2011).
56. Verleden, S. E., Sacreas, A., Vos, R., Vanaudenaerde, B. M. & Verleden, G. M. Advances in understanding bronchiolitis obliterans after lung transplantation. *Chest* **150**, 219–225 (2016).
57. Verleden, S. E. et al. The site and nature of airway obstruction after lung transplantation. *Am. J. Respir. Crit. Care Med.* **189**, 292–300

(2014).

58. Ghorani, V., Boskabady, M. H., Khazdair, M. R. & Kianmeher, M. Experimental animal models for COPD: a methodological review. *Tob. Induc. Dis.* **15**, 25 (2017).
59. Diamond, J. M. et al. Clinical risk factors for primary graft dysfunction after lung transplantation. *Am. J. Respir. Crit. Care Med.* **187**, 527–534 (2013).
60. Kiselev, V. Y. et al. SC3: consensus clustering of single-cell RNA-seq data. *Nat. Methods* **14**, 483–486 (2017).
61. Zappia, L. & Oshlack, A. Clustering trees: a visualization for evaluating clusterings at multiple resolutions. *Gigascience* **7**, giy083 (2018).
62. Van den Berge, K. et al. Trajectory-based differential expression analysis for single-cell sequencing data. *Nat. Commun.* **11**, 1201 (2020).
63. Wu, T. et al. clusterProfiler 4.0: a universal enrichment tool for interpreting omics data. *Innovation* **2**, 100141 (2021).
64. Gotoh, S. et al. Generation of alveolar epithelial spheroids via isolated progenitor cells from human pluripotent stem cells. *Stem Cell Rep.* **3**, 394–403 (2014).
65. Kathiriya, J. J. et al. Human alveolar type 2 epithelium transdifferentiates into metaplastic KRT5<sup>+</sup> basal cells. *Nat. Cell Biol.* **24**, 10–23 (2022).

## Acknowledgements

We thank the patients who contributed to our study, without their willingness to participate in research, these studies would not be possible; the staff at the Cell and Developmental Microscopy Core at the University of Pennsylvania, the Flow Cytometry Core at The Children's Hospital of

Philadelphia and the Animal Model Core of the Cystic Fibrosis Research Center of the University of Alabama at Birmingham for assistance in these studies; and M. Beers, V. Krymskya, S. Millar, A. Vaughan and S. Albelda for their critiques and insights throughout the development of the project. This work was supported by grants from the National Institutes of Health including HL148857, HL087825, HL134745 and HL132999 (to E.E.M.); 5T32HL007586-35 (to M.C.B.); 5R03HL135227-02 (to E.C.); K23 HL121406 (to J.M.D.); K08 HL150226 (to J.K.); DK047967, HL152960 and Federal Contract 75N92019R0014 (to J.F.E.); R35HL135816, P30DK072482, and U01HL152978 (to S.M.R.), R35HL150767 and U01HL134766 (to H.A.C.); and F32HL143931-01A1 and K99HL155785-01 (to J.J.K.). E.E.M. was also supported by the BREATH Consortium/Longfunds of the Netherlands. J.K. was supported by the Parker B. Francis Foundation. F.L.C.-D. was supported by a postdoctoral fellowship from GSK (RA3000034436). The embryonic stem cell experiments in this manuscript were not funded as part of the research agreement between University of Pennsylvania and GSK.

## Author information

### Author notes

1. These authors contributed equally: Maria C. Basil, Fabian L. Cardenas-Diaz

## Affiliations

1. Department of Medicine, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA

Maria C. Basil, Fabian L. Cardenas-Diaz, Michael P. Morley, Justine Carl, Jeremy Katzen, Katherine J. Slovik, Apoorva Babu, Su Zhou, Madison M. Kremp, Shanru Li, Joseph D. Planer, Yun Ying, Kathleen M. Stewart, Michelle Oyster, Jason D. Christie, Joshua M. Diamond & Edward E. Morrisey

2. Penn-CHOP Lung Biology Institute, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA

Maria C. Basil, Fabian L. Cardenas-Diaz, Michael P. Morley, Justine Carl, Jeremy Katzen, Katherine J. Slovik, Apoorva Babu, Su Zhou, Madison M. Kremp, Shanru Li, Joseph D. Planer, Rebecca Windmueller, Yun Ying, Kathleen M. Stewart, Jason D. Christie, Edward Cantu & Edward E. Morrisey

3. Department of Medicine, University of California, San Francisco, San Francisco, CA, USA

Jaymin J. Kathiriya, Alexis N. Brumwell & Harold A. Chapman

4. Cardiovascular Research Institute, University of California, San Francisco, San Francisco, CA, USA

Harold A. Chapman

5. Penn Cardiovascular Institute, University of Pennsylvania, Philadelphia, PA, USA

Michael P. Morley, Katherine J. Slovik & Apoorva Babu

6. Center for Regenerative Medicine, Boston University and Boston Medical Center, Boston, MA, USA

Katherine B. McCauley & Darrell N. Kotton

7. Department of Medicine and the Gregory Fleming James Cystic Fibrosis Research Center, University of Alabama at Birmingham, Birmingham, AL, USA

Shah S. Hussain & Steven M. Rowe

8. Department of Anatomy and Cell Biology, Carver College of Medicine, University of Iowa, Iowa City, IA, USA

Xiaoming Liu & John F. Engelhardt

9. Division of Cardiovascular Surgery, Department of Surgery, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA

Edward Cantu

10. The Pulmonary Center and Department of Medicine, Boston University and Boston Medical Center, Boston, MA, USA

Darrell N. Kotton

11. Department of Cell and Developmental Biology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA

Rebecca Windmueller & Edward E. Morrisey

## Contributions

M.C.B., F.L.C.-D., J.J.K., J.C., A.N.B., J.K., K.J.S., S.Z., M.M.K., S.L., J.D.P., S.S.H., Y.Y. and K.M.S. performed human primary cell experiments. K.J.S., K.B.M. and M.C.B. performed the human stem cell experiments. M.C.B., F.L.C.-D., J.J.K., M.P.M., A.B., R.W., D.N.K. and E.E.M. analysed the data. M.C.B., J.K., M.O., J.D.C., J.M.D. and E.C. provided access to human samples. X.L., S.S.H., J.F.E. and S.M.R. provided access to ferret samples. M.C.B. and E.E.M. wrote the first draft of the manuscript. All of the authors contributed to writing the final manuscript.

## Corresponding author

Correspondence to [Edward E. Morrisey](#).

## Ethics declarations

## Competing interests

The authors declare no competing interests.

# Peer review

## Peer review information

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

## Additional information

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## Extended data figures and tables

### [Extended Data Fig. 1 All sequenced patients contributed to all observed clusters in normal human epithelium and clusters are identified based on expression of known markers.](#)

**A)** Distribution of individual patient data is shown overlaying concatenated UMAP of the full distal data set (top row) and the subset of epithelium (bottom row). **B)** The patient contribution to each of the epithelial clusters is shown, demonstrating that each patient contributed to each epithelial population. RAS cell cluster marked with red asterisk. **C)** The UMAPs of each patient that contributed to the proximal data sets is shown for the full data set (top row) and the subset of epithelium (bottom row). **D)** The patient contribution to each epithelial cluster is shown in stacked bar graph. **E)** Output of clustree analysis as used to determine optimal cluster resolution. Green box denotes chosen resolution for epithelial analysis shown, red box denotes RAS cell cluster, yellow box denotes Secretory cell cluster. Gene expression between canonical Secretory cells and RAS cells were compared and **F)** GO analysis and **G)** WikiPathway analysis was performed on differentially expressed genes between RAS cells and Secretory cells. Shown are categories either up or down regulated in RAS cells as compared to Secretory cells. Previously identified cell lineage markers were examined to identify various cell types within the epithelium in distal (**H, I**) and

proximal (**J**, **K**) epithelial subsets. **H, J**) Feature plots show graded gene expression and distribution in UMAP. **I,K**) Violin plots show selected gene expression per cluster.

**Extended Data Fig. 2 Epithelial clusters within the normal distal human lung each harbor unique gene expression patterns.**

**A)** Heatmap of top cluster defining genes within the distal human lung epithelium demonstrates that *SCGB3A2*+ RAS cells have a distinct transcriptional signature. Of note, AT2\_c is a very small cluster with a small subset of cluster defining genes that are expressed at low levels and are not identified in most other cells. **B)** IHC of *SCGB3A2*, *LAMP3*, and *RAGE* on human lung distal parenchyma, demonstrating the interdigitation between respiratory bronchioles and alveolar units. N = 3. Scale bar = 100  $\mu$ m. **C)** In contrast to the distal human lung in B, distal mouse lung shows the termination of the mouse terminal bronchiole directly into the alveolar space. N = 2. Scale bar = 50  $\mu$ m. **D)** Separate channels from IHC presented in Fig. 1h are presented to allow for enhanced evaluation of each protein.

**Extended Data Fig. 3 Epithelial clusters in the normal human lung proximal airways lack the distinct secretory subtypes found in distal airways.**

**A)** Heatmap of cluster defining genes within the proximal human epithelium. **B)** Selected gene expression compared across all proximal epithelial clusters shown as dot plot. **C)** RNAscope analysis for *SCGB1A1* and *SCGB3A2* transcripts reveals a similar distribution to protein expression in distal airways; A distal airway transitioning from terminal to respiratory bronchiole is shown, highlighting the preserved gradient of *SCGB1A1* expressing cells in larger airways to *SCGB3A2* expressing cells in more distal airways. Scale bar = 500  $\mu$ m. N = 3. **D, E)** Additional airway structure showing intermixing of *SCGB1A1* and *SCGB3A2* positive populations. Yellow arrows point to double positive cells along gradient. While with the scRNA analysis there is *SCGB1A1* expression in *SCGB3A2* positive cells, double positive cells were rarely observed in RNAscope or protein IHC,

likely due to a threshold of the technique. Yellow arrows indicate rare double positive cells in a distal human respiratory airway. Red arrows indicate *SCGB3A2* single positive cells. Scale bars = 50  $\mu$ m. N = 3. **F, G**) IHC of *SCGB1A1* and *SCGB3A2* expression in large and distal airways from representative individual patients (n = 6). Scale bar = 100  $\mu$ m.

**Extended Data Fig. 4 Integration of proximal and distal single cell transcriptomes defines the unique gene signature and location of human RASCs.**

**A)** Distal and proximal scRNA-sequencing results from the same patient were concatenated, epithelium was isolated and re-clustered, and cell types were identified based on known markers. **B)** Distal and Proximal epithelium contribution to the resultant UMAP. **C)** Expression of *SCGB1A1* and *SCGB3A2*. Secretory cell cluster outlined in yellow, RAS cell clustered outlined in Red. **D)** Violin plots showing expression of selected genes, highlighting unique genes for each cluster, and shared gene expression between secretory and RAS cell clusters (orange box) and RAS cell and AT2 cell clusters (green box). **E)** Feature plots of both the proximal and distal epithelium highlighting key gene expression patterns used to identify known cell subtypes. **F)** Contribution of proximal and distal samples to each cell type, demonstrating that some groups are unique to each region while some are shared. **G)** Heat map showing gene expression of Secretory, RAS cell, and AT2 cell gene signatures from proximal and distal concatenated data set. Of note, these pairwise comparisons were done between the displayed groups only, as such the resultant gene expression is distinct from the cluster-defining genes we observed in Extended Data Fig. [2a](#), where the comparisons were done between all epithelial populations. Origin of cells shown in blue (proximal) and orange (distal) bar across top of heat map.

**Extended Data Fig. 5 Trajectory of primary human proximal and distal epithelium demonstrates both anticipated and novel epithelial progenitor relationships.**

**A)** Proximal airways epithelium clustering and resultant cell populations shown for reference for following analysis. **B)** Trajectory analysis in the proximal human airways results in 3 putative lineage relationships. UMAP plots show trajectories, color represents cell localization along pseudotime. Heatmaps showing expression of trajectory defining genes across pseudotime. **C)** Trajectory analysis of scRNAseq gene expression across distal epithelial populations suggests several epithelial relationship, including **D)** a putative relationship between RAS cells and AT2 cells. **E)** Heatmap of defining gene expression changes along pseudotemporal ordering of cells from RAS cell-to-AT2 cell trajectory. Pseudotemporal ordering shown in bar above heatmap. **F)** Additional epithelial trajectories were identified by slingshot analysis within the distal epithelium. For the additional trajectories that are not a focus of this current study, the individual trajectory is shown along with the corresponding heatmap of trajectory defining genes. Top indicated position along trajectory as color coded by position of the cell along pseudotemporal ordering. **G)** Gene expression in UMI count of selected genes along pseudotemporal ordering of distal trajectory 1 (panel **d**, **e**) on the x-axis.

**Extended Data Fig. 6 Human ES cell model of RAS cells demonstrates capacity to differentiate into AT2-like cells *in vitro*.**

**A)** General experimental schematic of development of iRAS cells and propagation in airway or transition to alveolar media. **B)** Brightfield microscopy showing organoid formation and fluorescence after sorted iRAS cells were grown in airway (top) or alveolar (bottom) media. Scale bar = 100  $\mu$ m. **C)** Corresponding flow cytometry analysis of endogenous mCherry reporter. **D)** Organoids from 3D culture with EPCAM or SCGB3A2 and SFTPC staining demonstrating SFTPC expression in alveolar organoids (bottom) and retention of SCGB3A2 expression in airways organoids (top). Scale bar = 100  $\mu$ m, 20  $\mu$ M for enlarged regions. N = 3 for panels **B–D**. **E)** Schematic of SFTPC protein processing in AT2 cells and **F)** western blot of primary translation product and processing intermediates (top, bands 1 and 2) and mature S-PC protein (band 3) in NKX2.1 progenitors, iRAS cells propagated in airway media, iRAS cells grown in alveolar media, and primary human AT2 cells (HT2-280+ cells).

Molecular weights indicated on left. N = 3. **G**) SFTPC staining in individual iAT2 cells (top) and primary human AT2 cells (bottom) showing punctate nature of SFTPC staining. Scale bar 10  $\mu$ m. N = 4. **H**) q-RT-PCR from iRASC grown in airway (red) or alveolar (green) media demonstrates differential gene expression of known RAS cell and AT2 cell marker genes. Data presented as box with median (bar) and upper and lower quartiles (box bounds) and whiskers for min and max values (n = 5). Unpaired Student's test performed \*p < 0.05, \*\*p < 0.01. **I**) Gene expression based on scRNA sequencing of iRAS cells, and iRAS cells after 14 days in alveolar media showing downregulation of primary RAS cell defining genes and upregulation of AT2 defining genes. All N refer to biological replicates

.

**Extended Data Fig. 7 Reference based integration of primary epithelium and hES cell derived epithelial populations.**

**A)** UMAP analysis showing the distribution of primary human epithelial populations within the concatenated data set of primary human epithelium, iRAS cells, and iAT2 cells. Grey cells represent non-primary human cells (ES derived populations). Colors correspond to primary human cell populations as indicated. All primary epithelial populations are shown. **B)** UMAP showing distribution of the hES cell populations included in the concatenated data set. Colors correspond to hES cell populations based on gene expression of *SCGB3A2* and *SFTPC* as indicated. **C)** Percentage of hES cells within primary cell defined clusters within concatenated data set. Clusters were identified based on localization of primary human epithelium. **D)** Venn diagram of Transcription factors identified as upregulated in primary RAS and iRAS cells compared to primary AT2 and iAT2 cells, respectively. **E)** Expression of Notch pathway genes *HES1* and *HES4* in primary RAS and iRAS cells compared to AT2 cell counterparts. **F)** Reference based integration of primary adult human epithelium, fetal human lung epithelial from day 11.5, 15, 18, and 21 (the days which included *SCGB3A2*+ secretory cell progenitor populations) from published fetal lung data set<sup>32</sup>, and iRAS cells. All three *SCGB3A2*+ populations clustered together, and *SCGB3A2*+ cells are shown in red. **G)** *SCGB3A2*+

cells were selected and re-clustered. **H**) A stacked bar graph of the contribution of each population to each resultant cluster.

**Extended Data Fig. 8 Dual reporter system demonstrates SFTPC expression in iAT2 cells and highlights dynamics of cell transitions *in vitro*.**

**A)** Schematic of vectors for SFTPC-eGFP targeting in the dual reporter hES cell line. **B)** Brightfield microscopy showing mCherry and eGFP expression in 3D cultures of sorted iRAS cells grown in either airway (top) or alveolar (bottom) media for 14 days. Scale bar = 100  $\mu$ m. N = 3. **C)** Flow cytometry showing endogenous SCGB3A2-mCherry and SFTPC-eGFP expression in iRAS cells grown in airway (top) or alveolar (bottom) media compared to NKX2.1 progenitor controls. **D)** Gating strategy for all hES cell flow cytometry and FACS experiments shown in Extended Data Figs. [6c](#), [8b](#), and [9h, i and n](#). **E)** Flow cytometry corresponding to experiment in Fig. [2a](#), of SCGB3A2-mCherry and SFTPC-eGFP expression over time as iRASC were propagated in Alveolar media. N = 3. **F)** UMAP of scRNA-seq analysis of all populations derived from iRAS cells at day 14 of differentiation in alveolar media reveals several clusters. **G)** The resulting culture was heterogenous and included both iAT2 lung endoderm progenitors as well as a small number of other foregut endoderm cell types. Feature plots showing expression of canonical AT1 and AT2 cell alveolar epithelial markers, Airway cell markers, Lung endoderm progenitor markers, neuroendocrine and tuft cell markers, and gastric fate markers, allowing putative identification of all observed clusters.

**Extended Data Fig. 9 Transition of iRAS cells to iAT2 cells is similar to the primary RAS cell-AT2 transition and is partially regulated by Notch and Wnt.**

**A)** Schematic of time-course of scRNA-seq experiment. **B)** Integration of entire time course showing cell origin, cell-cycle phase, and gene expression of *SCGB3A2* and *SFTPC*. **C)** Clustering of complete time course from iRAS to iAT2 cells development shows multiple clusters within the various time points. **D)** Trajectory analysis showed multiple putative

pseudotemporal orderings (top), and selected curve for further analysis based on termination in day 14 non-mitotic iAT2 cells (bottom). **E**) Heatmap of iRAS cell to iAT2 cell trajectory displaying genes defining the primary RAS to AT2 cell transition from Extended Data Fig. [5e](#). **F**) Expression of a subset of genes identified in primary RAS to AT2 cell transition shown over pseudotime in iRAS to iAT2 cell transition. **G**) Canonical airway and alveolar epithelial marker genes expression within the resultant UMAP. **H**) Flow cytometry of mCherry and eGFP expression in iRAS cells grown in Airway media in the presence or absence of DAPT, and corresponding percent fold change in mCherry expression and MFI. N = 4. **I**) Flow cytometry of mCherry and eGFP expression of iRAS cells grown in Alveolar media in the presence or absence of DAPT showing percent change in eGFP expression and MFI. **J**) q-RT-PCR of bulk populations from iRAS cells grown in airway media with or without DAPT. N = 3. **K**) q-RT-PCR from iRAS cells grown in alveolar media with or without DAPT. N = 3. Quantification of flow cytometry analysis of **L**) mCherry and **M**) eGFP positive single cells in culture after iRAS cells were grown in either Airway media or Airway media supplemented with CHIR99021. Data are represented as mean +/- SD and unpaired two-tailed t-tests performed. \*p < 0.05, \*\*p < 0.01. **N**) Representative flow cytometry plots (representative of n = 4). **O**) The heat map of each observed trajectory in panel d is presented with the top trajectory defining genes identified, and cell are ordered by pseudotemporal order on x-axis. All N represent biological replicates

[Source data](#)

[Extended Data Fig. 10 Identification and isolation of CEACAM6+ distal lung epithelial cells and demonstration that SCGB3A2+ cells can be identified by CEACAM6 and isolated from distal lung parenchyma.](#)

**A)** scRNA sequencing was performed on flow sorted Epcam+ HT2-280<sup>neg</sup> cells in order to enrich for epithelial cells of the human distal airways. **B)** UMAP of re-clustering of selected secretory airway cell populations demonstrate refined heterogeneity in *SCGB3A2* expressing cells. **C)** View

of distal airway showing distribution of *SCGB3A2*, *CEACAM6*, and *SFTPC* expression and **D**) Zoom in of region highlighting *SCGB3A2* and *CEACAM6* staining. White arrows indicate double positive cells. Of note, there are a small minority of *CEACAM*<sup>+</sup>/*SCGB3A2*<sup>low</sup> cells present in some sections. Scale bars = 100  $\mu\text{m}$ , 50  $\mu\text{m}$  for enlarged regions. N = 5. **E**) Representative cytospins and associated **F**) quantification of *CEACAM6*<sup>+</sup> population for pro-SFTPC reveals that <5% of cells are positive for this canonical AT2 cell marker (n = 3), compared to HT2-280<sup>+</sup> cells where over 90% of cells are positive for Pro-SFTPC (n = 2). Data presented as mean +/- SD. White arrow indicates rare SFTPC<sup>+</sup> cell in *CEACAM*<sup>+</sup> population. Scale bar = 50  $\mu\text{m}$ . **G**) Feature plots demonstrates that surface marker *CEACAM6* expression overlaps *SCGB3A2* expression. **H**) FACS approach to isolate *CEACAM6*<sup>+</sup>/HT2-280<sup>-</sup>/NGFR<sup>-</sup> airway cells. HT2-280 neg population was sorted for a *CEACAM6*<sup>+</sup> and NGFR<sup>-</sup> population. **I**) Gating strategy for obtaining populations in panel **H**. After gating on single cells, immune cells, endothelial cells and dead cells were excluded prior to selecting EPCAM<sup>+</sup> cells for further subsetting as shown in **H**

[Source data](#)

[Extended Data Fig. 11 All COPD patients contributed to all resultant clusters and gene expression within the COPD epithelium highlights the various epithelial cell clusters.](#)

Each patient contribution is shown overlying the entire concatenated UMAP for **A**) the entire data set (top row), and the subset of epithelial cells (bottom row). **B**) Stacked bar graph showing the patient contribution to each epithelial cell cluster, demonstrating that all patients contributed to all resultant clusters. RAS cell cluster marked with red asterisk. **C**) Canonical marker genes used to identify epithelial clusters are shown. **D**) Violin plots demonstrate distribution of known canonical marker genes across the various epithelial populations. **E**) Cluster defining genes within the COPD epithelium are shown in dot plot format. There are distinct changes in the transcriptome of non-epithelial populations in COPD. GO analysis of inter-cluster gene expression comparing select endothelial (**F**) and mesenchymal (**G**) populations suggests that the differential gene expression seen in the epithelium is distinct to that population.

## Extended Data Fig. 12 Epithelial cell gene expression differences in disease and the RAS cell to AT2 cell transition is altered in COPD.

**A)** Concatenation of normal and COPD peripheral samples and subset of epithelium showing expected epithelial populations. **B)** Identification of RAS cell and SCGB3A2+ AT2 populations based on expression of markers indicated. **C)** GO analysis of intra-cluster gene expression of RAS cells comparing COPD and healthy patient derived cells. **D)** Violin plots of selected genes contributing to GO processes in **(C)**. **E)** GO analysis of inter-cluster gene expression comparing SCGB3A2+ AT2 cells from COPD donors and AT2 cells from healthy donors and **F)** corresponding selected gene expression. **G)** GO analysis of intra-cluster gene expression of AT2 cells from COPD and healthy controls. **H)** Violin plots of selected genes from GO processes in **(G)**. For all, up regulated is COPD compared to healthy controls. **I)** Transcriptional inference analysis of the concatenated data set revealing multiple trajectories initiating at RAS cells. **J, K)** Comparison of gene expression along trajectory 1 (T1) versus trajectory 2 (T2) demonstrating differential gene expression changes along pseudotemporal ordering between the RAS to AT2 cell trajectory and the RAS to SCGB3A2+ AT2 cell trajectory. **L)** Distribution of individual patient data is shown overlaying concatenated UMAP of the COPD and healthy peripheral data sets. All cells are shown in top two rows, and epithelial subsets in bottom rows. **M)** Stacked bar graphs highlight patient level contribution to each cluster. RAS cells are indicated in red asterisk.

## Supplementary information

### Supplementary Fig. 1

Source gels. Source gels for Extended Data Fig. 6f. Molecular mass values are reprinted adjacent to the panels.

### Reporting Summary

### Supplementary Table 1

Patient characteristics. A list of all of the patients who were included in the analyses presented herein. Age, gender, self-identified race, and cause of death or disease at time of transplantation are provided. Where available, smoking history, FEV1, and/or arterial partial pressure of O<sub>2</sub> to fraction of inspired O<sub>2</sub> (*P/F* ratio) is also reported. The use of tissue is indicated by the type of experiment.

### **Supplementary Table 2**

Transcription factor analysis in RAS cells. Transcription factors upregulated in the hES cell model system (iRAS cells) compared with iAT2 cells and transcription factors upregulated in primary RAS cells compared with in primary AT2 cells are listed by log-transformed fold change between RAS cell and AT2 cell populations, with adjusted *P* values. We used MAST, which uses generalized linear model to perform univariate differential expression analysis, and *P* values were determined using a two-sided test. Multiple-testing correction was performed using the Benjamini–Hochberg procedure.

### **Supplemental Table 3**

RT–qPCR primers. A list of all of the RT–qPCR primers used in the included experiments listed.

## **Source data**

### **Source Data Fig. 3**

### **Source Data Fig. 4**

### **Source Data Extended Data Fig. 6**

### **Source Data Extended Data Fig. 9**

### **Source Data Extended Data Fig. 10**

# Rights and permissions

[Reprints and Permissions](#)

## About this article

### Cite this article

Basil, M.C., Cardenas-Diaz, F.L., Kathiriya, J.J. *et al.* Human distal airways contain a multipotent secretory cell that can regenerate alveoli. *Nature* **604**, 120–126 (2022). <https://doi.org/10.1038/s41586-022-04552-0>

- Received: 03 February 2021
- Accepted: 14 February 2022
- Published: 30 March 2022
- Issue Date: 07 April 2022
- DOI: <https://doi.org/10.1038/s41586-022-04552-0>

### Share this article

Anyone you share the following link with will be able to read this content:

[Get shareable link](#)

Sorry, a shareable link is not currently available for this article.

[Copy to clipboard](#)

Provided by the Springer Nature SharedIt content-sharing initiative

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

- Article
- [Published: 30 March 2022](#)

# Plant phytochrome B is an asymmetric dimer with unique signalling potential

- [Hua Li<sup>1</sup> ✉](#),
- [E. Sethe Burgie<sup>2</sup> ✉](#),
- [Zachary T. K. Gannam<sup>2</sup>](#),
- [Huilin Li](#) ORCID: orcid.org/0000-0001-8085-8928<sup>1</sup> &
- [Richard D. Vierstra](#) ORCID: orcid.org/0000-0003-0210-3516<sup>2</sup>

*Nature* volume **604**, pages 127–133 (2022)

- 3054 Accesses
- 185 Altmetric
- [Metrics details](#)

## Subjects

- [Cryoelectron microscopy](#)
- [Light responses](#)

## Abstract

Many aspects of plant photoperception are mediated by the phytochrome (Phy) family of bilin-containing photoreceptors that reversibly interconvert between inactive Pr and active Pfr conformers<sup>1,2</sup>. Despite extensive

biochemical studies, full understanding of plant Phy signalling has remained unclear due to the absence of relevant 3D models. Here we report a cryo-electron microscopy structure of *Arabidopsis* PhyB in the Pr state that reveals a topologically complex dimeric organization that is substantially distinct from its prokaryotic relatives. Instead of an anticipated parallel architecture, the C-terminal histidine-kinase-related domains (HKRDs) associate head-to-head, whereas the N-terminal photosensory regions associate head-to-tail to form a parallelogram-shaped platform with near two-fold symmetry. The platform is internally linked by the second of two internal Per/Arnt/Sim domains that binds to the photosensory module of the opposing protomer and a preceding ‘modulator’ loop that assembles tightly with the photosensory module of its own protomer. Both connections accelerate the thermal reversion of Pfr back to Pr, consistent with an inverse relationship between dimer assembly and Pfr stability. Lopsided contacts between the HKRDs and the platform create profound asymmetry to PhyB that might imbue distinct signalling potentials to the protomers. We propose that this unique structural dynamism creates an extensive photostate-sensitive surface for conformation-dependent interactions between plant Phy photoreceptors and their signalling partners.

This is a preview of subscription content

## Access options

Subscribe to Nature+

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

\$29.99

monthly

[Subscribe](#)

Subscribe to Journal

Get full journal access for 1 year

\$199.00

only \$3.90 per issue

[Subscribe](#)

All prices are NET prices.

VAT will be added later in the checkout.

Tax calculation will be finalised during checkout.

Buy article

Get time limited or full article access on ReadCube.

\$32.00

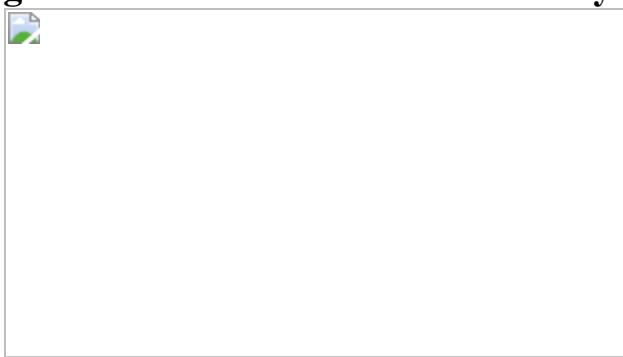
[Buy](#)

All prices are NET prices.

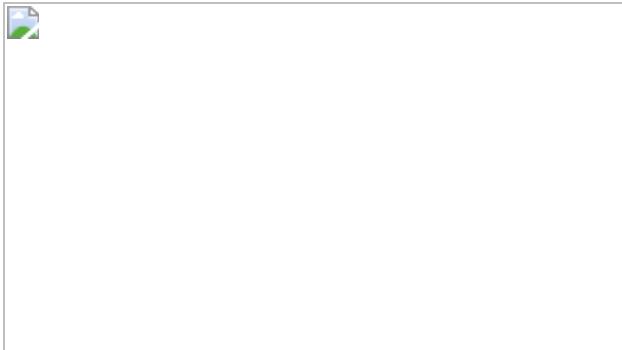
### **Additional access options:**

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

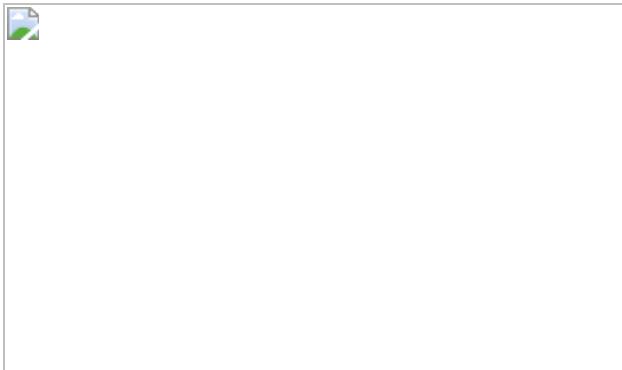
**Fig. 1: Overall 3D structure of the asymmetric *Arabidopsis* PhyB dimer.**



**Fig. 2: Asymmetry in the PhyB dimer.**



**Fig. 3:** The PSMs of PhyB are held by cross-protomer contacts between the PAS2 domain and the nPAS–GAF region that also destabilize Pfr.



**Fig. 4:** Interactions between the modulator loop and the PAS2 and PHY domains within each protomer stabilize PhyB dimerization but destabilize Pfr.



## Data availability

Full versions of all SDS–PAGE gels and blots are provided in Supplementary Fig. 1. The 3D cryo-EM map of the full-length *Arabidopsis* PhyB at 3.3 Å resolution has been deposited in the Electron Microscopy Data Bank database under accession code [EMD-24780](#). The corresponding atomic model has been deposited in the RCSB Protein Data Bank under accession code [7RZW](#). This study made use of several publicly available protein structures obtained from the RCSB Protein Data Bank (<http://www.rcsb.org>) under accession codes [4OUR](#), [6TC5](#), [3DGE](#), [4GCZ](#), [4U7O](#) and [4I5S](#). [Source data](#) are provided with this paper.

## References

1. Legris, M., Ince, Y. C. & Fankhauser, C. Molecular mechanisms underlying phytochrome-controlled morphogenesis in plants. *Nat. Commun.* **10**, 5219 (2019).
2. Burgie, E. S. & Vierstra, R. D. Phytochromes: an atomic perspective on photoactivation and signaling. *Plant Cell* **26**, 4568–4583 (2014).
3. Auldrige, M. E. & Forest, K. T. Bacterial phytochromes: more than meets the light. *Crit. Rev. Biochem. Mol. Biol.* **46**, 67–88 (2011).
4. Rockwell, N. C., Su, Y. S. & Lagarias, J. C. Phytochrome structure and signaling mechanisms. *Annu. Rev. Plant Biol.* **57**, 837–858 (2006).
5. Franklin, K. A. & Quail, P. H. Phytochrome functions in *Arabidopsis* development. *J. Exp. Bot.* **61**, 11–24 (2010).
6. Jung, J. H. et al. Phytochromes function as thermosensors in *Arabidopsis*. *Science* **354**, 886–889 (2016).
7. Legris, M. et al. Phytochrome B integrates light and temperature signals in *Arabidopsis*. *Science* **354**, 897–900 (2016).
8. Burgie, E. S. et al. Differing biophysical properties underpin the unique signaling potentials within the plant phytochrome families. *Proc. Natl Acad. Sci. USA* **118**, e2105649118 (2021).

9. Essen, L. O., Mailliet, J. & Hughes, J. The structure of a complete phytochrome sensory module in the Pr ground state. *Proc. Natl Acad. Sci. USA* **105**, 14709–14714 (2008).
10. Wagner, J. R., Brunzelle, J. S., Forest, K. T. & Vierstra, R. D. A light-sensing knot revealed by the structure of the chromophore-binding domain of phytochrome. *Nature* **438**, 325–331 (2005).
11. Yang, X., Kuk, J. & Moffat, K. Crystal structure of *Pseudomonas aeruginosa* bacteriophytochrome: photoconversion and signal transduction. *Proc. Natl Acad. Sci. USA* **105**, 14715–14720 (2008).
12. Takala, H. et al. Signal amplification and transduction in phytochrome photosensors. *Nature* **509**, 245–248 (2014).
13. Burgie, E. S., Zhang, J. & Vierstra, R. D. Crystal structure of *Deinococcus* phytochrome in the photoactivated state reveals a cascade of structural rearrangements during photoconversion. *Structure* **24**, 448–457 (2016).
14. Burgie, E. S. et al. Photoreversible interconversion of a phytochrome photosensory module in the crystalline state. *Proc. Natl Acad. Sci. USA* **117**, 300–307 (2020).
15. Isaksson, L. et al. Signaling mechanism of phytochromes in solution. *Structure* **29**, 151–160 (2021).
16. Anders, K., Daminelli-Widany, G., Mroginski, M. A., von Stetten, D. & Essen, L. O. Structure of the cyanobacterial phytochrome 2 photosensor implies a tryptophan switch for phytochrome signaling. *J. Biol. Chem.* **288**, 35714–35725 (2013).
17. Bhoo, S. H., Davis, S. J., Walker, J., Karniol, B. & Vierstra, R. D. Bacteriophytochromes are photochromic histidine kinases using a biliverdin chromophore. *Nature* **414**, 776–779 (2001).
18. Yeh, K. C., Wu, S. H., Murphy, J. T. & Lagarias, J. C. A cyanobacterial phytochrome two-component light sensory system.

*Science* **277**, 1505–1508 (1997).

19. Li, F. W. et al. Phytochrome diversity in green plants and the origin of canonical plant phytochromes. *Nat. Commun.* **6**, 7852 (2015).
20. Rockwell, N. C. & Lagarias, J. C. Phytochrome evolution in 3D: deletion, duplication, and diversification. *New Phytol.* **225**, 2283–2300 (2020).
21. Yeh, K. C. & Lagarias, J. C. Eukaryotic phytochromes: light-regulated serine/threonine protein kinases with histidine kinase ancestry. *Proc. Natl Acad. Sci. USA* **95**, 13976–13981 (1998).
22. Boylan, M. T. & Quail, P. H. Are the phytochromes protein kinases? *Protoplasma* **195**, 12–17 (1996).
23. Elich, T. D. & Chory, J. Phytochrome: if it looks and smells like a histidine kinase, is it a histidine kinase? *Cell* **91**, 713–716 (1997).
24. Ni, W. et al. A mutually assured destruction mechanism attenuates light signaling in *Arabidopsis*. *Science* **344**, 1160–1164 (2014).
25. Buckley, C. E. et al. Reversible optogenetic control of subcellular protein localization in a live vertebrate embryo. *Dev. Cell* **36**, 117–126 (2016).
26. Chernov, K. G., Redchuk, T. A., Omelina, E. S. & Verkhusha, V. V. Near-infrared fluorescent proteins, biosensors, and optogenetic tools engineered from phytochromes. *Chem. Rev.* **117**, 6423–6446 (2017).
27. Levskaya, A., Weiner, O. D., Lim, W. A. & Voigt, C. A. Spatiotemporal control of cell signalling using a light-switchable protein interaction. *Nature* **461**, 997–1001 (2009).
28. Shimizu-Sato, S., Huq, E., Tepperman, J. M. & Quail, P. H. A light-switchable gene promoter system. *Nat. Biotechnol.* **20**, 1041–1044 (2002).

29. Krall, L. & Reed, J. W. The histidine kinase-related domain participates in phytochrome B function but is dispensable. *Proc. Natl Acad. Sci. USA* **97**, 8169–8174 (2000).
30. Matsushita, T., Mochizuki, N. & Nagatani, A. Dimers of the N-terminal domain of phytochrome B are functional in the nucleus. *Nature* **424**, 571–574 (2003).
31. Burgie, E. S. et al. Photosensing and thermosensing by phytochrome B require both proximal and distal allosteric features within the dimeric photoreceptor. *Sci. Rep.* **7**, 13648 (2017).
32. Burgie, E. S., Bussell, A. N., Walker, J. M., Dubiel, K. & Vierstra, R. D. Crystal structure of the photosensing module from a red/far-red light-absorbing plant phytochrome. *Proc. Natl Acad. Sci. USA* **111**, 10179–10184 (2014).
33. Nagano, S. et al. Structural insights into photoactivation and signalling in plant phytochromes. *Nat. Plants* **6**, 581–588 (2020).
34. Diensthuber, R. P., Bommer, M., Gleichmann, T. & Moglich, A. Full-length structure of a sensor histidine kinase pinpoints coaxial coiled coils as signal transducers and modulators. *Structure* **21**, 1127–1136 (2013).
35. Wang, C. et al. Mechanistic insights revealed by the crystal structure of a histidine kinase with signal transducer and sensor domains. *PLoS Biol.* **11**, e1001493 (2013).
36. Cai, Y. et al. Conformational dynamics of the essential sensor histidine kinase WalK. *Acta Crystallogr. D* **73**, 793–803 (2017).
37. Casino, P., Rubio, V. & Marina, A. Structural insight into partner specificity and phosphoryl transfer in two-component signal transduction. *Cell* **139**, 325–336 (2009).
38. Lagarias, J. C. & Mercurio, F. M. Structure function studies on phytochrome. Identification of light-induced conformational changes

- in 124-kDa *Avena* phytochrome *in vitro*. *J. Biol. Chem.* **260**, 2415–2423 (1985).
39. Jones, A. M., Vierstra, R. D., Daniels, S. M. & Quail, P. The role of separate molecular domains in the structure of phytochrome from etiolated *Avena sativa* L. *Planta* **164**, 501–516 (1985).
  40. Li, H., Zhang, J., Vierstra, R. D. & Li, H. Quaternary organization of a phytochrome dimer as revealed by cryoelectron microscopy. *Proc. Natl Acad. Sci. USA* **107**, 10872–10877 (2010).
  41. Gourinchas, G. et al. Long-range allosteric signaling in red light-regulated diguanylyl cyclases. *Sci. Adv.* **3**, e1602498 (2017).
  42. Etzl, S., Lindner, R., Nelson, M. D. & Winkler, A. Structure-guided design and functional characterization of an artificial red light-regulated guanylate/adenylate cyclase for optogenetic applications. *J. Biol. Chem.* **293**, 9078–9089 (2018).
  43. Mechaly, A. E., Sassoon, N., Betton, J. M. & Alzari, P. M. Segmental helical motions and dynamical asymmetry modulate histidine kinase autophosphorylation. *PLoS Biol.* **12**, e1001776 (2014).
  44. Shin, A. Y. et al. Evidence that phytochrome functions as a protein kinase in plant light signalling. *Nat. Commun.* **7**, 11545 (2016).
  45. Klose, C., Nagy, F. & Schafer, E. Thermal reversion of plant phytochromes. *Mol. Plant* **13**, 386–397 (2020).
  46. Klose, C. et al. Systematic analysis of how phytochrome B dimerization determines its specificity. *Nat. Plants* **1**, 15090 (2015).
  47. Rensing, S. A., Sheerin, D. J. & Hiltbrunner, A. Phytochromes: more than meets the eye. *Trends Plant Sci.* **21**, 543–546 (2016).
  48. Kikis, E. A., Oka, Y., Hudson, M. E., Nagatani, A. & Quail, P. H. Residues clustered in the light-sensing knot of phytochrome B are

- necessary for conformer-specific binding to signaling partner PIF3. *PLoS Genet.* **5**, e1000352 (2009).
49. Ni, W. et al. PPKs mediate direct signal transfer from phytochrome photoreceptors to transcription factor PIF3. *Nat. Commun.* **8**, 15236 (2017).

## Acknowledgements

Cryo-EM data were collected using the Titan Krios system at the David Van Andel Cryo-Electron Microscopy Suite at the Van Andel Institute. We thank G. Zhao and X. Meng for help with data collection; J. Zhang for the mass spectrometry analysis; H. Zaher for help with the kinase assays; and C. Sherman and K. McLoughlin for technical assistance. This work was funded by the US National Institutes of Health R01 grants GM127892 (to R.D.V.) and GM131754 (to Huilin Li), and funds provided by the Van Andel Institute (to Huilin Li) and Washington University in St Louis (to R.D.V.).

## Author information

### Author notes

1. These authors contributed equally: Hua Li, E. Sethe Burgie

## Affiliations

1. Department of Structural Biology, Van Andel Institute, Grand Rapids, MI, USA

Hua Li & Huilin Li

2. Department of Biology, Washington University in St Louis, St Louis, MO, USA

E. Sethe Burgie, Zachary T. K. Gannam & Richard D. Vierstra

## Contributions

Hua Li, E.S.B., Huilin Li, Z.T.K.G. and R.D.V. designed the experiments. Hua Li performed the cryo-EM and 3D reconstruction. Hua Li and E.S.B. built and refined the atomic models. E.S.B. and Z.T.K.G. performed the mutagenesis and spectroscopy assays. Z.T.K.G. conducted the limited proteolysis and kinase assays. Hua Li, E.S.B., Huilin Li and R.D.V. wrote the manuscript with input from Z.T.K.G.

## Corresponding authors

Correspondence to [Huilin Li](#) or [Richard D. Vierstra](#).

## Ethics declarations

### Competing interests

The authors declare no competing interests.

### Peer review

### Peer review information

*Nature* thanks Jorge Casal, Elizabeth Getzoff and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

## Additional information

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## Extended data figures and tables

## Extended Data Fig. 1 Workflow and resolution estimation for the cryo-EM map of *Arabidopsis* PhyB.

**a**, SDS-PAGE analysis of the recombinant full-length biliprotein. Gels were either stained for protein with Coomassie blue (left) or assayed for bound PΦB by zinc-induced fluorescence (right). MM, molecular mass standards. Samples were indistinguishable to those described by Burgie et al.<sup>2</sup> **b**, UV-vis absorbance spectra of PhyB. The spectra were collected from dark-adapted samples (Pr) or after saturating irradiation with 660-nm red light (RL, mostly Pfr). Absorption maxima were determined from the difference spectrum shown at 70% amplitude. The spectral change ratio (SCR)<sup>8</sup> at 723 nm is indicated in parenthesis. **c**, Work flow for data processing of the cryo-EM images of the PhyB dimer. In the first refined overall map at 3.4-Å resolution, all PhyB domains were present but the regions encompassing the PAS1 domain were poorly resolved. Focused refinement, excluding the PAS1 domains, led to the 3.3-Å final map (lower left panel). Lower right panel shows that the EM density of the flexible PAS1 domains (purple), which were captured at 15-Å resolution by 3D variability analysis of down-sampled particle images. **d–h**, Resolution estimation of the 3.3-Å 3D map. **d**, A representative cryo-EM micrograph sampled from 6,153 micrographs collected. **e**, Sampling of 2D class averages. **f**, Colored-coded local resolution of the 3D map. **g**, Eulerian angle distribution of raw particle images used in the final 3D reconstruction. **h**, Gold-standard Fourier shell correlation (FSC) and the validation of the atomic model by correlation curves comparing the model to the final and two half maps.

## Extended Data Fig. 2 Superposition of the cryo-EM map densities of the PSM, PΦB, and the bilin-binding GAF domain pocket with the X-ray crystallographic model of the PSM.

Motifs/residues are colored as in Fig. 1b. **a**, Fitting of the nPAS, GAF and PHY domains, and the hairpin (HP) motif within the EM map density (light grey surface) of protomer A and protomer B. PΦB is shown in red. **b**, Superposition of the PhyB PSM determined by cryo-EM of the full-length PhyB (protomer A; slate blue) and by X-ray crystallography of the PhyB PSM (grey; residues 90-624, PDB ID code, 4OUR<sup>32</sup>). **c**, PΦB

conformations (in sticks) in protomers A and B modeled within the EM map density (grey mesh). The A-D pyrrole rings are labeled along with Cys357 that forms a thioether linkage to the C3<sup>1</sup> carbon of PΦB. The D ring C18<sup>2</sup> carbon is indicated. **d**, Superposition of the PΦB structures determined by cryo-EM of the full-length PhyB protomer A (colored) and by X-ray crystallography of the PhyB PSM (grey). **e**, The bilin-binding pockets of protomers A and B highlighting neighboring amino acids (sticks) and superposed in the EM map density (grey mesh). **f**, Superposition of the bilin-binding pocket determined from the cryo-EM structure of full-length PhyB protomer A with the X-ray crystallographic structure of the PhyB PSM.

**Extended Data Fig. 3 Superposition of the cryo-EM map densities of various PhyB motifs with the cryo-EM model.**

Shown in sticks are various amino acids modeled within the EM densities (grey mesh). Domains/residues are colored as in Fig. 1b. **a–c**, The NTE and knot lasso region of protomer A showing the knot lasso (a) and NTE separately (b), and combined (c). **d**, The hairpin loop extending from the PHY domain in protomer A. **e**, Portions of the DHp regions within the HKRD from protomers A and B. Gln937, which positionally corresponds to the conserved phosphoacceptor histidine found in prokaryotic transmitter histidine kinases, is circled in red. **f**, The modulator loop extending from between the PAS1 and PAS2 domains within protomer A. **g**, Residues within the CA domain of the HKRD in protomer A surrounding a possible ATP-binding pocket.

**Extended Data Fig. 4 Topology of PhyB generated from the cryo-EM 3D model and structural predictions of the three PAS domains.**

**a**, Topological schematic of PhyB. Shown are the NTE, nPAS, GAF, PHY and PAS2 domains, and the DHp and CA domains within the HKRD highlighting the positions, lengths, and contacts of the α-helices and β-strands, and the contacts for the knot lasso, hairpin, and modulator loop features. The position of PΦB within the GAF domain β-sheet is indicated.

The entry and exit points of the poorly resolved PAS1 domain between the PHY and PAS2 domains are shown. The GAF and nPAS loops unique to plant Phys and the PHY domain hairpin residues (WGG and PRXSF) involved in a predicted  $\beta$ -strand to  $\alpha$ -helical transformation during photoconversion are identified in the ellipsoids<sup>32,33</sup>. The predicted ATP-binding region is highlighted by the yellow ovals. The H1a cruciate region within the helix  $\alpha$ 1 of the DHp, which provides the head-to-tail to head-to-head crossover point with Cys925 at its center, is located by the dark blue box. Gln937 is the residue that replaces the phosphoacceptor histidine found in prokaryotic two-component HKs. Amino acid sequence conservation of each feature can be found in Extended Data Fig. 6. **b-d**, Structural predictions of the three PAS domains in PhyB using TrRosetta<sup>62</sup>. The PAS domain cores are circled by the dashed grey line, which is followed by an  $\alpha/\beta$  roll<sup>64</sup>. Terminal amino acids are indicated. **b** and **c**, Superposition of 3D models of the nPAS and PAS2 domains determined by cryo-EM with those predicted (p) by TrRosetta. The cartoons on the left are the cryo-EM models and those on the right are superpositions of the models (grey) with those calculated (rainbow) (rmsd = 1.1 Å for nPAS and 1.0 Å for PAS2). **d**, Predicted models of the PAS1 region by TrRosetta. The left cartoon is a prediction for the PAS1 domain plus 31 additional N-terminal residues not found in the cryo-EM model. The middle cartoon includes the top five predictions for the PAS1 domain alone. The right cartoon is a superposition of the predicted PAS1 domain with the PAS2 cryo-EM model (rmsd = 1.3 Å).

### [Extended Data Fig. 5 Limited protease sensitivity is consistent with the cryo-EM model of PhyB.](#)

**a**, Concentration-dependent cleavage of PhyB by chymotrypsin and GluC. Purified full-length PhyB was incubated for 15 min with increasing amounts of each protease and then subjected to complete digestion with trypsin followed by tandem MS identification of peptides generated by each protease. Peptides that ended in chymotrypsin or GluC cut sites were quantified from the MS1 scans. Each row represents a potential cleavage site; white bars indicate no cleavage whereas green boxes represent regions without detectable peptides. All MS data represent the means of four technical replicates. The digestions were aligned with the domain

architecture of PhyB (see Fig. 1a). **b**, Relative susceptibility to proteolysis for all cleaved sites at 2 ng/μL chymotrypsin or GluC. Bars are colored by domain as in (a). **c**, Proportion of cleavage sites within each domain that were either susceptible to high or low concentrations of both proteases, not cleaved, or not detected. **d–f**, 3D views of the protease-sensitive sites in PhyB highlighting: (d) the NTE-nPAS-GAF-PHY-PAS2, (e) HKRD, and (f) PSM regions. The PhyB structure is shown in cartoon while the cleavage sites are highlighted in spheres and color-coded based on protease sensitivity. Protomers A and B are presented in grey and white, respectively. Residues involved in dimerization are highlighted in magenta in (f).

### Extended Data Fig. 6 Amino acid sequence alignment of the PhyA, PhyB, PhyC and PhyE subfamilies within angiosperms.

See ref. <sup>8</sup> for full description of the sequence list. The font height of each amino acid is proportional to its percent homology within each Phy isoform subfamily. The positions of the NTE, nPAS, GAF, PHY, PAS1, PAS2, and DHp and CA domains of the HKRD are located by the red, blue, green, orange, gray, magenta, brown and cyan bars, respectively. The PAS and GAF loops, the knot lasso, and the hairpin and modulator features are located by light blue, light green, turquoise, yellow, and dark red bars, respectively. The α-helices and β-strands, along with their numbering within each domain, are shown below the sequence by the coiled and wavy lines, respectively. The red star locates the position that commonly contains the phosphoacceptor histidine within prokaryotic transmitter HK domains. The blue circles indicate the core amino acids within the N-terminal knot. Red arrows locate the end point for the six N-terminal truncations of PhyB analyzed in this study (N624, N778, N799, N908, N928, and N982). Green arrowheads locate the residues shown experimentally to promote GAF-PAS2 contacts within the dimer. Red diamonds locate residues predicted to form the ATP-binding pocket in the CA domain based on prokaryotic transmitter HKs. A green circle locates Cys925 that is at the center of the cruciate crossover that transitions the PhyB protomers from head-to-tail to head-to-head arrangements.

## Extended Data Fig. 7 Structural and enzymatic analyses of PhyB reveal its homology to transmitter HKs but with a compromised phosphotransferase activity.

**a**, Cartoon 3D structure of the paired HKRDs from *Arabidopsis* (*At*) PhyB showing the structures and inter-molecular interfaces between the CA and DHp domains. Images on the right show a pair of orthogonal views with residues within one half of the HKRD dimer interface shown as spheres. These residues were contributed by helix  $\alpha$ 1 from the DHp of protomer B, and helix  $\alpha$ 2 of DHp and helices  $\alpha$ 1,  $\alpha$ 2 and  $\alpha$ 4 from the CA domain of protomer A. **b**, The network of intermolecular contacts between the DHp and CA domains in (a) illustrated for simplicity. **c**, Top views of the DHp regions of the HKRDs for *At* PhyB as compared to the same region in the prokaryotic HK853 transmitter HK from *Thermotoga maritima* (PDB ID 3DGE<sup>37</sup>). Gln937 in *At* PhyB and the phosphoacceptor histidine in *Tm* HK853 are shown in red sticks. **d**, Closeup 3D views of the DHp domains in *At* PhyB corresponding to the region surrounding phosphoacceptor histidine in transmitter HKs. Gln937 in PhyB, which is a histidine in transmitter kinases, is circled. **e**, 3D superposition of the CA domain in *At* PhyB shown in cartoon with those from several bacterial two-component HKs illustrating its HK ancestry. Representatives include YF1 from *Bacillus subtilis* (*Bs*) (PDB ID 4GCZ), WalK from *Lactobacillus plantarum* (*Lp*) (PDB ID 4U7O), HK853 from *Thermotoga maritima* (*Tm*) (PDB ID 3DGE), and Vick from *Streptococcus mutans* (*Sm*) (PDB ID 4I5S). **f**, Model showing the predicted position of ADP (red) in the *At* PhyB CA domain when modeled after that for *Lp* WalK. Residues that might participate in binding are indicated. ADP clashes with multiple residues in the pocket of this predicted *At* PhyB model, suggesting that conformational shifts in *At* PhyB induced by ATP or upon photoactivation would be necessary for binding. **g**, Schematic of binding interactions between the ADP analogue adenylyl-imidodiphosphate (AMPPNP) and CA domain from the *Lp* WalK determined by X-ray crystallography (left; PDB ID 4U7O<sup>36</sup>) and that predicted for *At* PhyB when modelled after the *Lp* WalK structure (right). Hydrogen bonds and representative hydrophobic interactions are indicated with green and red dashed lines, respectively. Analogous residues are depicted in similar positions in schematics, except

for *LpAsn514* and *AtSer1054*. **h** and **i**, *AtPhyB* is a poor protein kinase as compared to *Pseudomonas syringae* (*Ps*) BphP based on autophosphorylation assays. The recombinant biliproteins were incubated at ambient temperature (~24°C) with 150 μM ATP supplemented with 10 μCi of [ $\gamma$ -<sup>32</sup>P]-ATP, quenched with SDS-PAGE sample buffer, and subjected to SDS-PAGE. Shown are the SDS-PAGE gels assayed for bound <sup>32</sup>P by autoradiography or stained for protein with Coomassie blue. **h**, Time course for autophosphorylation of *PsBphP* as Pfr. **i**, Comparisons of autophosphorylation activities of *AtPhyB* as Pr and Pfr with those of *PsBphP*. Reactions containing equal mass amounts of biliprotein were terminated after 2 hr. (left) Autoradiography of the kinase reactions. (right) SDS-PAGE gel showing the biliprotein preparations used. Arrowheads locate *PsBphP*. The phosphorimager scans are representative of 3 independent experiments. Full gels can be found in [Supplementary Fig. 1](#).

**Extended Data Fig. 8 SDS-PAGE analyses and absorption spectra of the PhyB truncations and point mutations studied here.**

**a** and **d**, UV-vis absorbance spectra. The absorption spectra were collected from dark-adapted samples (Pr) or after saturating irradiation with 660-nm red light (RL, mostly Pfr). Absorption maxima were determined from the difference spectrum shown at 70% amplitude. SCR values are indicated in parentheses. Spectra represent the mean of three technical replicates. **b** and **c**, SDS-PAGE analysis of the purified PhyB proteins described in (a) and (d). After electrophoresis, the biliproteins were either stained for protein with Coomassie blue or imaged for the bound PΦB by zinc-induced fluorescence under UV light. Full gels can be found in [Supplementary Fig. 1](#).

**Extended Data Table 1 Data collection, processing, model refinement, and validation statistics for the *Arabidopsis* PhyB dimer**

**Extended Data Table 2 Comparisons of sister domains in PhyB by superposition of all matching C $\alpha$  atoms**

**Extended Data Table 3 Thermal reversion rate constants for the *Arabidopsis* PhyB mutant collection**

# Supplementary information

## Supplementary Information

Supplementary Methods and the accompanying references, and the legends for Supplementary Fig. 1, Supplementary Videos 1 and 2 and the source data.

## Reporting Summary

## Supplementary Fig. 1

## Supplementary Video 1

Cryo-EM structure of *Arabidopsis* PhyB

## Supplementary Video 1

Morph between protomers A and B of *Arabidopsis* PhyB

## Source data

## Source Data Fig. 3

## Source Data Fig. 4

## Rights and permissions

## Reprints and Permissions

## About this article

## Cite this article

Li, H., Burgie, E.S., Gannam, Z.T.K. *et al.* Plant phytochrome B is an asymmetric dimer with unique signalling potential. *Nature* **604**, 127–133 (2022). <https://doi.org/10.1038/s41586-022-04529-z>

- Received: 01 September 2021
- Accepted: 07 February 2022
- Published: 30 March 2022
- Issue Date: 07 April 2022
- DOI: <https://doi.org/10.1038/s41586-022-04529-z>

## Share this article

Anyone you share the following link with will be able to read this content:

[Get shareable link](#)

Sorry, a shareable link is not currently available for this article.

[Copy to clipboard](#)

Provided by the Springer Nature SharedIt content-sharing initiative

---

This article was downloaded by **calibre** from <https://www.nature.com/articles/s41586-022-04529-z>

[Download PDF](#)

- Article
- [Published: 07 February 2022](#)

# Pyrimidine inhibitors synergize with nucleoside analogues to block SARS-CoV-2

- [David C. Schultz](#) [ORCID: orcid.org/0000-0002-7890-8815<sup>1</sup>](#),
- [Robert M. Johnson](#) [ORCID: orcid.org/0000-0002-1976-7688<sup>2 na1</sup>](#),
- [Kasirajan Ayyanathan<sup>3 na1</sup>](#),
- [Jesse Miller](#) [ORCID: orcid.org/0000-0002-7990-4912<sup>3 na1</sup>](#),
- [Kanupriya Whig](#) [ORCID: orcid.org/0000-0001-9114-4365<sup>1</sup>](#),
- [Brinda Kamalia<sup>1</sup>](#),
- [Mark Dittmar](#) [ORCID: orcid.org/0000-0003-2884-5630<sup>3</sup>](#),
- [Stuart Weston<sup>2</sup>](#),
- [Holly L. Hammond<sup>2</sup>](#),
- [Carly Dillen<sup>2</sup>](#),
- [Jeremy Ardanuy<sup>2</sup>](#),
- [Louis Taylor<sup>2</sup>](#),
- [Jae Seung Lee<sup>3</sup>](#),
- [Minghua Li<sup>3</sup>](#),
- [Emily Lee<sup>4</sup>](#),
- [Clarissa Shoffler](#) [ORCID: orcid.org/0000-0002-9064-2378<sup>5</sup>](#),
- [Christopher Petucci<sup>5</sup>](#),
- [Samuel Constant<sup>6</sup>](#),
- [Marc Ferrer<sup>4</sup>](#),
- [Christoph A. Thaiss](#) [ORCID: orcid.org/0000-0001-8226-7718<sup>7</sup>](#),
- [Matthew B. Frieman](#) [ORCID: orcid.org/0000-0003-0107-0775<sup>2</sup>](#) &
- [Sara Cherry](#) [ORCID: orcid.org/0000-0003-3956-6610<sup>1,3,7</sup>](#)

*Nature* volume 604, pages 134–140 (2022)

- 16k Accesses

- 4 Citations
- 225 Altmetric
- [Metrics details](#)

## Subjects

- [Drug screening](#)
- [SARS-CoV-2](#)
- [Virology](#)

## Abstract

The SARS-CoV-2 virus has infected more than 261 million people and has led to more than 5 million deaths in the past year and a half<sup>1</sup> (<https://www.who.org/>). Individuals with SARS-CoV-2 infection typically develop mild-to-severe flu-like symptoms, whereas infection of a subset of individuals leads to severe-to-fatal clinical outcomes<sup>2</sup>. Although vaccines have been rapidly developed to combat SARS-CoV-2, there has been a dearth of antiviral therapeutics. There is an urgent need for therapeutics, which has been amplified by the emerging threats of variants that may evade vaccines. Large-scale efforts are underway to identify antiviral drugs. Here we screened approximately 18,000 drugs for antiviral activity using live virus infection in human respiratory cells and validated 122 drugs with antiviral activity and selectivity against SARS-CoV-2. Among these candidates are 16 nucleoside analogues, the largest category of clinically used antivirals. This included the antivirals remdesivir and molnupiravir, which have been approved for use in COVID-19. RNA viruses rely on a high supply of nucleoside triphosphates from the host to efficiently replicate, and we identified a panel of host nucleoside biosynthesis inhibitors as antiviral. Moreover, we found that combining pyrimidine biosynthesis inhibitors with antiviral nucleoside analogues synergistically inhibits SARS-CoV-2 infection in vitro and in vivo against emerging strains of SARS-CoV-2, suggesting a clinical path forward.

## Main

SARS-CoV-2 is a coronavirus, which is a family of single-stranded positive-sense RNA viruses, at least seven of which infect humans. RNA viruses including coronaviruses replicate using a virally encoded RNA-dependent RNA polymerase (RdRp), and nucleoside analogues, which are incorporated by the RdRp into the growing viral RNA chain, are a large class of approved direct-acting antivirals<sup>3</sup>. Depending on the analogue, incorporation can lead to chain termination or

mutagenesis, ultimately inhibiting viral replication<sup>4</sup>. RdRps have conserved structures, and thus nucleoside analogues can show broad activity across related and unrelated viruses<sup>5,6</sup>. Therefore, repurposing efforts have identified nucleoside analogues that are active against newly emerging viruses, and such efforts have discovered that the nucleoside analogue remdesivir inhibits SARS-CoV-2 replication, becoming the first approved antiviral therapeutic against this novel coronavirus<sup>7,8</sup>.

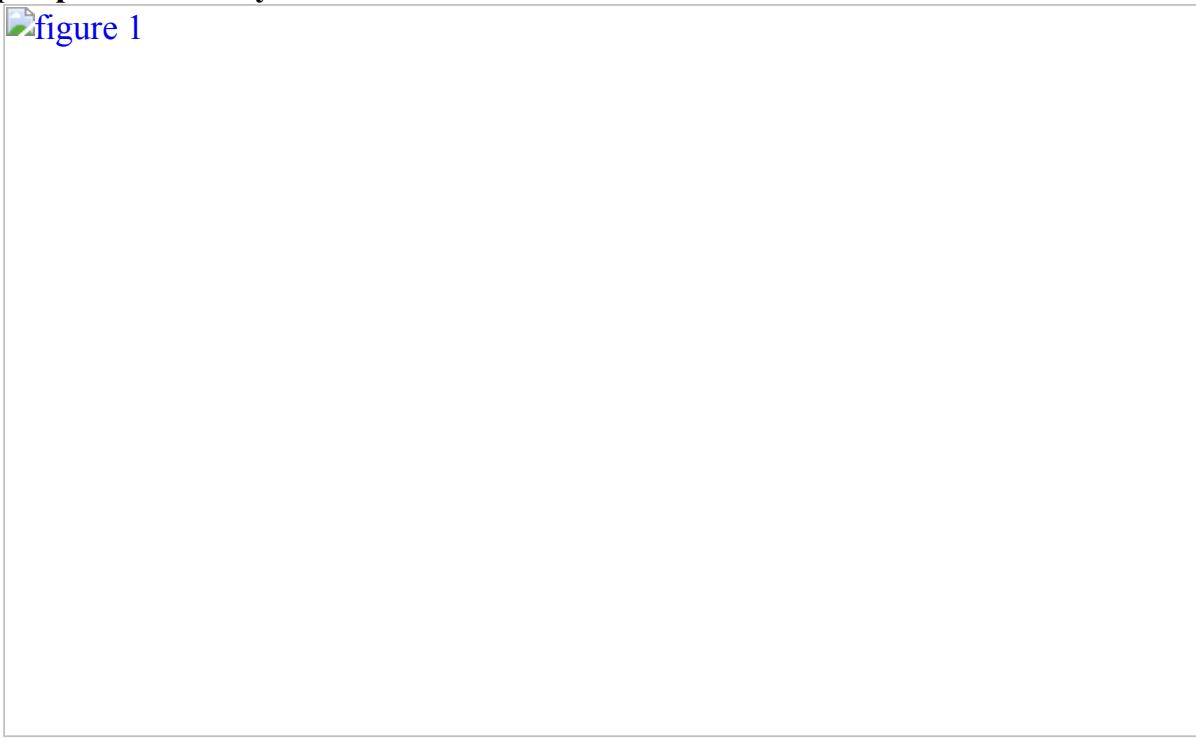
As all viruses, including SARS-CoV-2, are dependent on diverse cellular factors and metabolic products for their replication, the identification of host-directed antivirals also shows promise. In particular, host nucleoside biogenesis is required for viral replication as RNA viruses require high levels of nucleoside triphosphates for their growth. Widespread efforts are underway to identify essential host pathways that are druggable, and to repurpose therapeutics against these host targets. Respiratory epithelial cells are the major cellular target for SARS-CoV-2 *in vivo*. We and others have found that the cellular entry pathway as well as other host-dependent steps in the viral lifecycle show cell-type-specific differences and we were most interested in those antivirals that would be active in the respiratory tract<sup>9</sup>. Therefore, we used the human respiratory cell line Calu-3 for these studies. We set out to screen small-molecule libraries that contained approved drugs, drugs in clinical trials and drugs with known targets to uncover both direct-acting and host-directed antivirals using the ancestral SARS-CoV-2 virus (WA1) and a cell-based, high-content assay in respiratory cells. We optimized a microscopy-based assay to achieve robust screening parameters ( $Z' > 0.5$ ) using vehicle (DMSO) and remdesivir (10  $\mu$ M) as controls on each plate. We infected Calu-3 cells pre-treated with compounds and quantified infection of SARS-CoV-2 48 h after infection using an antibody to double-stranded RNA, a viral replication intermediate<sup>9</sup>. In addition, we quantified the number of cells in each well to remove drugs that were cytotoxic.

We screened approximately 18,000 drugs from three repurposing libraries: an in-house PENN library of approximately 3,500 drugs, about 3,400 drugs from the NCATS repurposing collection chosen to avoid overlap with the PENN library, and the ReFrame collection of roughly 11,300 drugs, most of which have been tested in humans<sup>10</sup>. Together, we screened a large fraction of the drugs that can potentially be rapidly repurposed, complementing previous screening efforts<sup>9,11,12</sup>. We screened the PENN library in duplicate at 0.8  $\mu$ M and identified 77 drugs that inhibited infection by more than 60% and that had little toxicity (more than 80% cell viability) in each screen; the NCATS library at 0.8  $\mu$ M and identified 45 drugs that inhibited infection by more than 60% and had little toxicity (more than 80% cell viability); and the ReFrame library in duplicate at 3.6  $\mu$ M and identified 135 compounds that inhibited infection by more than 60% and had low toxicity (more than 60% cell viability) in both replicates (Extended Data Fig. [1a–c](#)). We validated the candidates by repurchasing powders (PENN and NCATS) or testing pre-spotted validation plates

(ReFrame) followed by in-house-created dose–response studies. This allowed us to determine potency (half maximal effective concentrations ( $EC_{50}s$ )) and toxicity (half maximal cytotoxic concentrations ( $CC_{50}s$ )) of each compound and focus on the 122 non-redundant compounds that had a selective index (SI) greater than 3 in Calu-3 cells (Fig. 1a, Supplementary Table 1 (and 102 drugs had a  $SI > 10$ )). These 122 compounds fall into a number of general categories, with nucleoside analogues accounting for 13% of the validated candidates.

**Fig. 1: Antiviral nucleosides are highly active in respiratory cells and show cell-type-specific activity.**

 figure 1



**a**, Pie chart of antivirals validated in Calu-3 cells with selective index ( $EC_{50}/CC_{50} > 3$ ). From approximately 18,000 compounds, 122 show activity. **b**, Sixteen nucleoside analogues validated in Calu-3 cells with the nucleoside type listed along with the  $EC_{50}$ ,  $CC_{50}$  and SI. **c**, Calu-3, A549-ACE2 or Caco-2 cells were treated with the indicated nucleosides in dose–response showing infection (blue) and toxicity (green). Data are presented as mean  $\pm$  s.d. ( $n = 3$  independent biological replicates). BCNA, bicyclic nucleoside analogue. **d**, Calu-3 cells pretreated with vehicle or 10  $\mu M$  of the indicated drugs and infected with SARS-CoV-2 for 48 h and total RNA was subject to RT–qPCR analysis of viral infection. Data are presented as mean  $\pm$  s.e.m. for reduction compared to vehicle control ( $n = 4$  independent biological replicates) (\*\* $P < 0.001$ , one-way ANOVA; values are listed in Extended Data Fig. 1f). vRNA, viral RNA. **e**, RT–qPCR analysis of viral infection for the indicated drugs (remdesivir (10  $\mu M$ ), molnupiravir (10  $\mu M$ ), azathioprine (30  $\mu M$ ) and mercaptapurine (30  $\mu M$ )) in

nasal ALI cultures. Data are presented as mean  $\pm$  s.e.m. ( $n = 2$  independent biological replicates). **f**, BLISS analysis of the  $2 \times 2$  combination of remdesivir and EIDD-2801 in Calu-3 cells showing additivity. Data are presented as mean values of excess over BLISS for  $n = 9$  independent biological replicates. The statistical significance was determined by a one-sample Student's *t*-test (\* $P < 5 \times 10^{-2}$ ).

Nucleoside analogues are a common class of drugs that are synthetic analogues that mimic their physiological counterparts and can be incorporated by cellular polymerases into DNA or RNA to inhibit cell division<sup>4,13</sup>. In addition, these can act as antimetabolites that deplete the supply of deoxynucleotides needed for DNA replication or nucleotides for RNA synthesis through inhibition of nucleoside biosynthesis enzymes<sup>4</sup>. Nucleoside analogues and antimetabolites are generally used to treat cancer and for immunosuppression. A subgroup of nucleoside analogues are direct-acting antivirals as they are incorporated specifically by viral polymerases, leading to defects in viral nucleic acid replication. Additional nucleoside analogues have been shown to have antiviral activity, probably through their antimetabolite activity; they are not incorporated but rather compete with cellular enzymes for their natural ligands. Two direct-acting antiviral nucleosides have been shown to block SARS-CoV-2 infection in vitro and in animal models and have been approved for use in humans<sup>7,14</sup>. Remdesivir is an adenosine analogue incorporated by the viral RdRp leading to delayed chain termination<sup>15,16</sup>. Molnupiravir (also known as EIDD-2801) is a cytosine analogue incorporated by the viral RdRp leading to increased viral mutagenesis and antiviral activity<sup>14,17,18</sup>. We identified remdesivir, the active metabolite of remdesivir, Gs-441524, and molnupiravir as antiviral in our unbiased screens (Fig. 1a, b). We identified 13 additional nucleoside analogues with antiviral properties against SARS-CoV-2 (Fig. 1b, Extended Data Fig. 2). To determine the breadth of antiviral activity of the nucleosides, we tested a panel of cell lines that are permissive to infection with SARS-CoV-2, including human respiratory A549-expressing human ACE2, human intestinal epithelial Caco-2 cells, human hepatocyte Huh7.5 cells and African green monkey Vero cells (Fig. 1c, Extended Data Fig. 1d). The known direct-acting antivirals remdesivir and molnupiravir show activity across diverse cell types, with variable EC<sub>50</sub>s (Fig. 1a–c, Extended Data Fig. 1d). We also found that the active form of molnupiravir, EIDD-1931, was active across cell types (Extended Data Fig. 3a). We also validated the antiviral activity of remdesivir and molnupiravir in Calu-3 cells using a quantitative reverse-transcription PCR (RT-qPCR) assay and found that these nucleoside analogues had strong antiviral activity by this orthogonal assay (Fig. 1d).

Some of the additional antiviral nucleoside analogues that we identified are thought to act as antimetabolites by competing with the natural ligands that inhibit enzymatic function and are generally used for cancer or immunosuppression<sup>4</sup> (for example, 6-mercaptopurine; Extended Data Fig. 1e). We found that these drugs are non-toxic in

Calu-3 cells at the antiviral concentrations (Fig. 1a–c, Extended Data Fig. 1d, e), at least in part because these cells divide slowly. Moreover, we found that these antiviral nucleoside analogues show very divergent cell-type-specific activity and toxicities (Fig. 1a–d, Extended Data Fig. 1d, e). For example, tubercidin displays antiviral activity in Calu-3, Caco-2 and Huh7.5 cells with toxicity in A549-ACE2 and Vero cells. By contrast, thioguanine and 6-mercaptopurine are active in Calu-3 and A549-ACE2 cells but are not active in Caco-2 or Vero cells. We also monitored the antiviral activity of a subset of these nucleoside analogues in Calu-3 cells by RT-qPCR, and found that they all had significant antiviral activity (Fig. 1d). We also tested a subset of these antiviral nucleosides in primary human nasal air–liquid interface (ALI) cultures and found that only remdesivir and molnupiravir showed potent antiviral activity as measured by RT–qPCR (Fig. 1e). These studies demonstrated cell-type-specific antiviral activities of nucleoside analogues.

Given that the two direct-acting antiviral nucleosides are distinct nucleoside derivatives, with remdesivir, an adenosine analogue, and molnupiravir, a cytosine analogue, we tested whether the combination of these nucleosides would show antiviral synergy. Synergy is defined as having greater than an additive effect. Calu-3 cells were treated with six concentrations of each drug individually and in combination with each other, resulting in a matrix of 36 drug–x drug–y concentration pairs monitoring viral infection. We used BLISS analysis to determine whether the two drugs interact<sup>19</sup>. The null hypothesis in this model is that the drugs are additive, with a positive interaction leading to synergy and a negative interaction leading to antagonism. We found that the co-treatment of Calu-3 cells with remdesivir and molnupiravir during SARS-CoV-2 infection was additive (Fig. 1f, Extended Data Fig. 3b–d).

Antimetabolites are thought to act, at least in part, through inhibition of nucleoside biosynthesis by binding to metabolic enzymes and competing with the natural ligands<sup>4</sup>. Cells have two pathways for nucleoside biogenesis, de novo synthesis and salvage pathways that recycle purines and pyrimidines from degradation products. Although salvage pathways sustain cell viability, they cannot supply sufficient amounts of nucleoside triphosphates to allow for fast proliferation or support viral replication<sup>20</sup>. Therefore, in addition to applications in cancer, the inhibition of the de novo nucleotide biosynthesis constitutes a broad-spectrum antiviral strategy. In support of this, our screen identified a subset of known inhibitors of host-encoded nucleoside biosynthetic enzymes within the pyrimidine and purine biosynthetic pathways shown in Fig. 2a. This included three pyrimidine biosynthesis inhibitors: the two dihydroorotate dehydrogenase (DHODH) inhibitors BAY-2402234 and Brequinar, as well as the uridine monophosphate synthase (UMPS) inhibitor pyrazofurin. Dose–response studies showed that these drugs were active in Calu-3 cells with low toxicity (Fig. 2b). We did not identify the classical IMP dehydrogenase (IMPDH) inhibitors

ribavirin or mycophenolic acid in our screens, although ribavirin and mycophenolic acid did inhibit viral infection at higher doses in Calu-3 cells (Extended Data Fig. [4a](#)). We also tested the human respiratory cells A549-ACE2 and the panel of human cell lines that are permissive to SARS-CoV-2, and found that the DHODH inhibitors were active, whereas pyrazofurin had modest activity and AVN944 showed toxicity (Extended Data Fig. [4b,c](#)).

**Fig. 2: Inhibitors of host nucleoside metabolism are antiviral.**

 figure 2

**a**, Simplified schematic of nucleoside metabolism. Number of arrows indicates single versus multiple steps. CAD, carbamoyl-phosphate synthetase 2, aspartate transcarbamylase and dihydroorotate; GMPS, guanosine monophosphate synthetase; PRPP, phosphoribosyl diphosphate synthase. **b**, Dose–response analysis of the pyrimidine biosynthesis or purine biosynthesis inhibitors in Calu-3 cells. Infection (blue) and toxicity (green) are shown. Data are presented as mean  $\pm$  s.d. ( $n = 3$  independent biological replicates). **c**, Analysis of Calu-3 cells treated with the indicated inhibitor in the presence of increasing concentrations of the indicated nucleosides. Data are presented as mean  $\pm$  s.d. ( $n = 3$  independent biological replicates) for per cent infection. **d**, Calu-3 cells treated with the indicated inhibitor for 24 h and cell pellets subjected to liquid chromatography–mass spectrometry to quantify the nucleoside triphosphate levels. Mean  $\pm$  s.e.m. is shown ( $n = 6$  independent

biological replicates). \* $P < 0.05$ , \*\* $P < 0.01$ , \*\*\* $P < 0.001$ , one-way analysis of variance (ANOVA) with Dunnett's multiple comparisons test ( $P = 0.0177$  (ATP)  $P = 0.0045$  (GTP)  $P < 0.0001$  (CTP) and  $P = 0.0009$  (UTP)).

We confirmed that the antiviral activity of the pyrimidine biosynthesis inhibitors was through nucleoside metabolism as we could fully block the antiviral activity of BAY-2402234, Brequinar or pyrazofurin by treating Calu-3 cells with the pyrimidine nucleosides cytidine and/or uridine, but not the purine nucleosides adenosine or guanosine (Fig. 2c). IMPDH is required for de novo guanosine biosynthesis, and thus treatment of cells with the purine guanosine, but not adenosine or the pyrimidine nucleosides, could reverse the antiviral activity of AVN944 (Fig. 2c). In addition, we monitored nucleoside pools upon treatment with the direct-acting nucleoside analogues remdesivir and molnupiravir and observed no changes in triphosphate levels of the four nucleotides (Fig. 2d). By contrast, treatment with the DHODH inhibitor Brequinar reduced levels of the pyrimidine triphosphates, whereas treatment with IMPDH inhibitor AVN944 reduced the levels of guanosine triphosphates with increases in the production of ATP, which is produced from a common precursor that is shunted into that pathway upon inhibition of IMPDH (Fig. 2d). These data demonstrate that altering nucleoside pools can block viral replication in the absence of cytotoxicity.

Next, we set out to determine whether altering pyrimidine or purine pools using host-directed metabolic inhibitors would synergize with the antiviral nucleoside analogues remdesivir or molnupiravir. Using our microscopy-based assay, we performed  $2 \times 2$  dose-response analysis. We found notable synergy between DHODH inhibitors (Brequinar or BAY-2402234) and nucleoside analogues (molnupiravir or remdesivir) (Fig. 3a, b, Extended Data Fig. 5). This synergy was observed in the submicromolar range of the DHODH inhibitor Brequinar and the low nanomolar range for BAY-2402234. We also tested whether the combination of molnupiravir or remdesivir with pyrazofurin, which inhibits UMPS, also shows synergy and found that it does (Extended Data Fig. 6a). By contrast, we observed little interaction between the IMPDH inhibitor AVN944 and either of the antiviral nucleoside analogues, with a modest trend towards antagonism (Fig. 3c, Extended Data Fig. 6b-d).

**Fig. 3: Combination of molnupiravir or remdesivir with DHODH inhibitors is synergistically antiviral in vitro.**

---

 **figure 3**

**a–c**, BLISS analysis in Calu-3 cells with molnupiravir or remdesivir in combination with the DHODH inhibitors Brequinar (**a**) or BAY-2402234 (**b**) and the IMPDH inhibitor AVN944 (**c**). Data are presented as mean values of excess over BLISS. The number of independent biological replicates (*n*) is indicated for the combination tested. Statistical significance was determined by a one-sample Student's *t*-test (\* $P < 5 \times 10^{-2}$ , \*\* $P < 10^{-3}$  and \*\*\* $P < 10^{-4}$ ). **d**, Calu-3 cells were treated with the indicated drugs and infected with SARS-CoV-2. Total RNA was subject to RT–qPCR 48 h after infection for analysis of infection. Data are presented as mean  $\pm$  s.e.m. for reduction compared to vehicle control (*n* = 3 independent biological replicates). The black asterisks show the *P* value relative to vehicle: \*\* $P < 0.01$ , using one-way ANOVA. The green asterisks show *P* values relative to single treatment with molnupiravir: \* $P < 0.05$ , \*\* $P < 0.01$  and \*\*\* $P < 0.001$ , using one-way ANOVA. *P* values are listed in Extended Data Fig. [8b](#). **e, f**, ALI-bronchial (**e**) or ALI-nasal (**f**) cells were treated with the indicated drugs and infected with SARS-CoV-2. Viral replication was quantified by RT–qPCR 72 h after infection and expression (vRNA/18S) was normalized to vehicle-treated cells. Data are presented as mean  $\pm$  s.e.m. (*n* = 2 independent biological replicates). In **d–f**, the indicated drug concentration is in  $\mu\text{M}$ .

Given the emergence of SARS-CoV-2 variants and the promise of antivirals to have activity against diverse strains of SARS-CoV-2, we also tested whether these drugs were active against SARS-CoV-2 variants (Alpha, Beta, Gamma and Delta) and found that all strains were sensitive to molnupiravir, remdesivir and the pyrimidine biosynthesis inhibitors (Extended Data Fig. [7a](#)). The ancestral strain (WA1) and Delta strain were also sensitive to AVN944, whereas the others were not. We focused on

molnupiravir as it is an orally bioavailable drug and thus can be potentially used in an outpatient setting and is currently undergoing approvals<sup>21,22</sup>. We tested whether the combination of molnupiravir and the DHODH inhibitors would show antiviral synergy against SARS-CoV-2 variants. We observed synergy between both DHODH inhibitors (Brequinar or BAY-2402234) and molnupiravir (Extended Data Fig. 7b). This suggests that limiting the pyrimidine pool in combination with the nucleoside analogue molnupiravir increases the antiviral activity of nucleoside analogues against diverse strains of SARS-CoV-2. The mechanism of this synergy is unclear; however, it may be due to increased uptake or incorporation of the nucleoside analogues when pyrimidines are limiting.

We further explored the activity and interactions between DHODH inhibitors and molnupiravir. We used an orthogonal assay to confirm that molnupiravir shows dose-dependent inhibition in Calu-3 cells using RT-qPCR (Fig. 3d, Extended Data Fig. 8a). As single agents, Brequinar and BAY-2402234 showed modest levels of inhibition as measured by RT-qPCR (Fig. 3d). However, when we combined treatments, we observed notable decreases in viral replication upon co-administration of either of these DHODH inhibitors and molnupiravir (Fig. 3d).

We next tested the activity of these inhibitors in air-liquid interface (ALI) respiratory epithelial cell cultures that more closely model the human respiratory epithelium. We used both tracheobronchial and nasal ALI cultures as these represent the two major sites of SARS-CoV-2 infection<sup>23</sup>. We performed additional studies to identify doses of these drugs that did not impact epithelial barrier function (by monitoring trans-epithelial electrical resistance), cilia beating frequency or toxicity (measuring lactate dehydrogenase release) in our nasal ALI cultures<sup>24</sup>. We found that treatment with molnupiravir, Brequinar or BAY-2402234 at concentrations up to 30 μM were non-toxic in this system (Extended Data Fig. 8c). Therefore, we used these doses as the maximum in our ALI cultures. In bronchial ALI cultures, we found that molnupiravir showed dose-dependent activity and that Brequinar had significant single-agent activity, whereas BAY-2402234 had little activity as a single agent in these cells (Fig. 3e). However, we found a significant reduction in viral replication upon co-treatment with either Brequinar or BAY-2402234 with molnupiravir (Fig. 3e). In the nasal cells, we found that molnupiravir showed dose-dependent activity and that neither Brequinar nor BAY-2402234 had significant single-agent activity (Fig. 3f). Again, we found a significant reduction in viral replication upon co-treatment of molnupiravir with either DHODH inhibitor (Fig. 3f).

Molnupiravir and Brequinar are both orally dosed drugs and molnupiravir has been approved in patients with COVID-19. We observed synergistic antiviral activity with this combination of drugs in diverse model cell systems; therefore, we set out to test whether this combination would also show benefit in the treatment of SARS-CoV-2

infection *in vivo*. We used a mouse model of infection in which wild-type BALB/C mice were intranasally inoculated with the SARS-CoV-2 Beta strain B.1.351 ( $1 \times 10^5$  plaque-forming units (p.f.u.) per mouse). In this model, we found robust replication 2 days after infection as measured by viral titres in the lungs (approximately  $10^8$  p.f.u per gram lung), bronchiolar sloughing of infected epithelial cells, and significant inflammatory cell infiltration including oedema with peribronchiolar and perivascular cuffing as measured by histology (Fig. 4). As expected, when we treated the animals with molnupiravir (EIDD-2801), we observed dose-dependent reduction in viral titres (Fig. 4a). In addition, there was only minor mitigation of the lung pathology with significant epithelial cell infection and inflammatory cell infiltration other than at the highest dose of molnupiravir tested ( $150 \text{ mg ml}^{-1}$ ) (Fig. 4b,c). We also treated mice with Brequinar alone, at doses similar to COVID-19 clinical trials (NCT04575038), and observed no effect on viral titres or lung pathology at 2 days after infection (Fig. 4a–c). Next, we combined treatments with multiple dosing of molnupiravir and a single dose of Brequinar ( $20 \text{ mg kg}^{-1}$ ). We found an approximately 4 log reduction in viral titres upon co-treatment with Brequinar and molnupiravir at the highest dose of molnupiravir, and a significant reduction in titres at both  $150 \text{ mg kg}^{-1}$  and  $50 \text{ mg kg}^{-1}$  molnupiravir when combined with Brequinar compared with molnupiravir alone (Fig. 4a). We observed strong suppression of inflammation in the lung where, even at the lowest combination dose, there was reduced peribronchiolar and perivascular cuffing with lessened alveolar and interstitial inflammation and oedema (Fig. 4b,c). Moreover, co-treatment led to a clear protection of lung architecture with little alterations to bronchiolar and alveolar cells. These low doses of Brequinar are outside the toxic range and show protection. We also performed an experiment in which we used a single dose of molnupiravir ( $50 \text{ mg kg}^{-1}$ ) alone, and in combination with increasing doses of Brequinar ( $20 \text{ mg kg}^{-1}$  and  $50 \text{ mg kg}^{-1}$ ) (Fig. 4d–f). We again observed significant decreases in viral titres upon co-treatment of molnupiravir and Brequinar over molnupiravir alone (Fig. 4d). Histological analysis revealed that treatment with molnupiravir modestly reduced SARS-CoV-2-induced pathology and that co-administration with Brequinar further reduced the inflammatory response seen with molnupiravir alone (Fig. 4b,c,e,f).

**Fig. 4: Combination of molnupiravir and Brequinar reduces SARS-CoV-2 infection and inflammation *in vivo*.**

---

 **figure 4**

---

Wild-type BALB/C mice were treated with Brequinar (intraperitoneal administration) and/or molnupiravir (oral administration) daily at the indicated concentrations starting 12 h before infection. Mice ( $n = 5$  per group over 2 independent experiments) were intranasally inoculated with  $1 \times 10^5$  p.f.u. per mouse of SARS-CoV-2 (B.1.351). **a–f**,

Lungs were analysed for viral titre 2 days after infection by plaque assay (**a, d**) or fixed in 4% paraformaldehyde for haematoxylin and eosin staining and quantified for interstitial inflammation (**b, c, e, f**).  $n = 5$  mice per group. Mean  $\pm$  s.d. is shown.  $*P < 0.05$ ,  $**P < 0.01$  and  $***P < 0.001$ , using non-parametric one-way ANOVA with Dunnett's multiple comparison test. The red asterisks are compared to vehicle; the blue asterisks are compared to molnupiravir.  $P$  values are listed in Extended Data Fig. [8d](#). Scale bars, 200  $\mu\text{M}$ .

### Source data

We also performed a study using therapeutic dosing in this model. We infected mice intranasally, and on day 1 after infection, we treated infected mice with molnupiravir, Brequinar or the combination and quantified viral infection and inflammation in the lungs on day 3 after infection. We found that the addition of Brequinar to molnupiravir reduced viral titres and inflammation (Extended Data Fig. [9](#)). Together, this combination of molnupiravir and Brequinar for COVID-19 treatment shows promise as we observed both reduction in viral replication and decreased pathology when dosed before or after infection.

Given that for many viral infections combinations of antivirals are needed to suppress infection, we suggest that combining nucleoside analogues with DHODH inhibitors would be beneficial, as the combination would both reduce replication and inflammation due to SARS-CoV-2 infection within therapeutic ranges. Owing to the recent success of the SARS-CoV-2 protease inhibitor Paxlovid (Pfizer; PF-07321332), we also tested the combination of this drug with molnupiravir or remdesivir against SARS-CoV-2 Beta or Delta strains. We found that these combinations are additive (Extended Data Fig. [10](#)). Thus, there are numerous potential combinations that should be tested in clinical trials that may alter the trajectory of the current pandemic.

## Methods

### Viruses and cells

SARS-CoV-2 WA1 was provided by the BEI/CDC. SARS-CoV-2 (Alpha, Beta, Gamma and Delta) was provided by A. Pekosz (Johns Hopkins University School of Public Health). Virus stocks were amplified using the ARTIC primer set<sup>[25](#)</sup> and sequenced using the MinION system (Oxford Nanopore Technologies) by the J. Craig Venter Institute (MD, USA) to more than 4,000 $\times$  genome coverage. Stock sequence was verified by aligning reads to the reference genome provided by the BEI (GSAID accession: EPI\_ISL\_890360) using minimap2 version 2.22 with the ‘map-ont’ presets<sup>[26](#)</sup>, followed by inspection of the consensus sequence and alignment using IGV<sup>[27](#)</sup>. Stocks had less than 1% variation. Stock virus was prepared by infection of

Vero E6 cells expressing TMPRSS2 in growth medium (DMEM (Quality Biological), supplemented with 10% (v/v) fetal bovine serum (Gibco), 1% (v/v) penicillin–streptomycin (Gemini Bio-products) and 1% (v/v) l-glutamine (2 mM final concentration; Gibco) fetal bovine serum plus for 2 or 3 days when cytopathic effect (CPE) was visible. Media were collected and clarified by centrifugation before being aliquoted for storage at –80 °C. Titre of stock was determined by plaque assay or 50% tissue culture infectious dose (TCID<sub>50</sub>) analysis using Vero E6 cells as previously described<sup>5</sup>. All work with infectious virus was performed in a biosafety level 3 laboratory and approved by the University of Pennsylvania Biosafety Committee and the University of Maryland School of Medicine Institutional Biosafety Committee.

Vero cells (American Type Culture Collection (ATCC), CCL81 or E6) and Huh7.5 cells (C. Rice, Rockefeller) were cultured in DMEM, supplemented with 10% (v/v) fetal bovine serum, 1% (v/v) penicillin–streptomycin and 1% (v/v) l-Glutamax (Invitrogen). Calu-3 cells (ATCC, HTB-55) were cultured in MEM, supplemented with 10% (v/v) fetal bovine serum, 1% (v/v) penicillin–streptomycin, 1% (v/v) l-glutamine and 1% (v/v) non-essential amino acids. A549-ACE2 cells were cultured in RPMI1640 supplemented with 10% (v/v) fetal bovine serum, 1% (v/v) penicillin–streptomycin and 1% (v/v) l-Glutamax. Caco-2 cells (ATCC, HTB-37) were cultured in MEM alpha supplemented with 20% (v/v) fetal bovine serum, 1% (v/v) penicillin–streptomycin and 1% (v/v) l-glutamine. All cells were grown at 37 °C, 5% CO<sub>2</sub> and 20% O<sub>2</sub>. All cell lines were validated by morphology and gene expression and are mycoplasma free.

Day 20 air–liquid interface (ALI) EpiAirway tracheobronchial tissues were obtained from MatTek (AIR-100) and used after day 30. Tissues were fed twice a week until use. ALI-pooled nasal epithelial cultures from Epithelix were fed twice a week until use. Toxicity studies were performed on nasal air–liquid interface cultures. Daily basolateral exposure of test compounds was evaluated using tissue integrity (trans-epithelial electrical resistance (TEER)); cytotoxicity was measured as basolateral lactate dehydrogenase (LDH) release and cilia beating frequency at 72 h.

### Tissue integrity (TEER)

Tissue integrity was determined by monitoring TEER using the EVOM3 Ohm Meter (World Precision Instruments). Resistance values ( $\Omega$ ) were converted to TEER ( $\Omega \text{ cm}^2$ ) using the following formula: TEER ( $\Omega \text{ cm}^2$ ) = (resistance value ( $\Omega$ ) – 100( $\Omega$ )) × 0.33 ( $\text{cm}^2$ ), where 100  $\Omega$  is the resistance of the membrane and 0.33  $\text{cm}^2$  is the total surface of the epithelium.

### Cytotoxicity (LDH release)

Basolateral LDH release was quantified using Cytotoxicity LDH Assay Kit-WST (Dojindo, CK12–20), measuring the absorbance of each sample at 490 nm with a microplate reader. To determine the percentage of cytotoxicity, the following equation was used ( $A$  = absorbance values): cytotoxicity (%) =  $(A \text{ (exp value)} - A \text{ (low control)}) / A \text{ (high control)} - A \text{ (low control)}$ )  $\times 100$ . The high-control value was obtained by 10% Triton X-100 apical treatment (24 h). Triton X-100 causes a massive LDH release and corresponds to 100% cytotoxicity.

### Cilia beating frequency

Cilia beating frequency was measured using the following setup: a Sony XCD V60 camera connected to an Olympus BX51 microscope with a  $\times 5$  objective and a camera-specific software. The cilia beating frequency is expressed as Hz. 256 images were captured at a high-frequency rate (125 frames per second) at 34 °C. Cilia beating frequency was then calculated using Cilia-X software (Epithelix).

### High-throughput screening

Ten-thousand Calu-3 cells were plated per well of 384-well assay plates (Corning) in 20  $\mu\text{l}$  of growth medium. For the PENN and NCATS libraries, 50 nl of drugs was added at a final concentration of 0.8  $\mu\text{M}$  in 0.2% DMSO. For the ReFrame library, 20  $\mu\text{l}$  of MEM was dispensed to compound source plates pre-spotted with 50 nl of 10 mM compound in DMSO. Of diluted compound, 5  $\mu\text{l}$  was added per well of 384-well assay plate, yielding a final concentration of 3.6  $\mu\text{M}$  in 0.04% DMSO. The positive control, 10  $\mu\text{M}$  remdesivir ( $n = 32$ ), and the negative control, 0.2% DMSO ( $n = 32$ ), were spotted on each plate. One hour after the addition of the drugs, cells were infected with SARS-CoV-2 (multiplicity of infection (MOI) = 0.5). Cells were fixed 40–48 h post-infection in 4% formaldehyde in PBS for 15 min at room temperature, washed three times with PBS, blocked with 2% BSA in PBST for 60 min, and incubated in primary antibody (anti-double-stranded RNA J2, absolute antibody, 1:500) overnight at 4 °C. Cells were washed three times in PBST with an automated plate washer (BioTek) and incubated in secondary antibody (anti-mouse Alexa 488, 1:1,000 and Hoechst 33342) for 1 h at room temperature. Cells were washed three times in PBST with an automated plate washer and imaged using an automated microscope (ImageXpress Micro, Molecular Devices). Cells were imaged with a  $\times 10$  objective, and four sites per well were captured. The total number of cells and the number of infected (double-stranded RNA $^+$ ) cells were measured using the cell scoring module (MetaXpress 5.3.3), and the percentage of infected cells was calculated. The aggregated percent infection of the 0.2% DMSO ( $n = 32$ ) and 10  $\mu\text{M}$  remdesivir control wells ( $n = 32$ ) on each assay plate were used to calculate  $z$ -factors, as a measure of assay performance and data quality. Sample well infection was normalized to aggregated DMSO plate control wells and expressed as percentage of control (POC = (%infection<sub>sample</sub>/average

$(\%infection_{DMSO}) \times 100$ ) and Z score ( $Z = (\%infection_{sample} - \text{average } \%infection_{DMSO})/\text{standard deviation } \%infection_{DMSO}$ ) in Spotfire (PerkinElmer). Candidate hits were selected as the following: for the UPENN screen, compounds with POC < 40% for infection and viability > 80% in either replicate; for the NCATS screen, POC < 40% for infection and viability > 80%; for the ReFrame screen, POC < 50% for infection and > 60% viability for the average of the replicates, compared to DMSO control.

## Dose-response studies

Candidate drugs from the UPENN and NCATS library were purchased as powders from Selleckchem, MedchemExpress, Cayman and MedKoo and suspended in DMSO. Drugs were arrayed in 8-pt dose-response in 384-well plates. For the ReFrame validation, 15  $\mu$ l of MEM was dispensed to compound source plates pre-spotted with 100 nl of compound in DMSO arrayed in an 8-pt dose-response. Of the diluted compound, 5  $\mu$ l was added per well of 384-well assay plate, yielding final concentrations of 9.5, 3.1, 1.0, 0.35, 0.12, 0.04, 0.01 and 0.004  $\mu$ M in 0.1% DMSO. Calu-3 ( $n = 10,000$ ), A549-ACE2 ( $n = 3,000$ ), Caco-2 ( $n = 1,500$ ), Huh7.5 ( $n = 3,000$ ) or Vero ( $n = 3,000$ ) cells were plated in 384-well plates. Twenty-four hours after plating (72 h for Caco-2), drug additions and infections were performed using the screening conditions. DMSO (0.2%;  $n = 32$ ) and 10  $\mu$ M remdesivir ( $n = 32$ ) were included on each plate as controls for normalization. For nucleoside rescue experiments, cells were treated with an EC<sub>90</sub> concentration of a given nucleoside biosynthesis inhibitor in combination with a dose-response of nucleosides. Infection at each drug concentration was normalized to aggregated DMSO plate control wells and expressed as POC (POC =  $\%infection_{sample}/\text{average } \%infection_{DMSO \ cont}$ )  $\times 100$ . A non-linear regression curve fit analysis (GraphPad Prism 8) was performed on POC of percent infection and cell viability using log<sub>10</sub>-transformed concentration values to calculate EC<sub>50</sub> values for percent infection and CC<sub>50</sub> values for cell viability for each drug-cell line combination. The EC<sub>50</sub> and CC<sub>50</sub> values represent the average of two or more independent experimental replicates. SI was calculated as a ratio of the CC<sub>50</sub> and EC<sub>50</sub> values (SI = CC<sub>50</sub>/EC<sub>50</sub>) of a drug.

## Quantitative drug combination analysis

Using our high-throughput screening assay conditions, drugs were added to Calu-3 cells in 384-well assay plates using a Tecan D300e dispenser in a matrix that combined six concentrations of remdesivir (up to 2  $\mu$ M) or molnupiravir (up to 10  $\mu$ M) with six concentrations of Brequinar (up to 10  $\mu$ M), BAY-2402234 (up to 0.5  $\mu$ M) and AVN944 (up to 2  $\mu$ M), resulting in 36 concentration combinations per drug

pair in 0.2% DMSO. Each combination was independently repeated at least three times as technical duplicates in each biological replicate. Sample well infection was normalized to aggregated 0.2% DMSO plate control wells ( $n = 32$ ) and expressed as POC. Synergy between drug combinations was determined by the BLISS independence model, to quantitatively assess drug interaction patterns within the drug–drug combination matrix .The BLISS expectation ( $E$ ) for a combined response was calculated by  $E = (A + B) - (A \times B)$  where  $A$  and  $B$  are the fractional inhibition of SARS-CoV-2 infection of drug A and drug B at a given dose. The difference between the BLISS expectation and the observed inhibition of SARS-CoV-2 infection for the combination of drug A and drug B at the same dose is the BLISS value. BLISS values between 0 and 10 indicate that the combination is additive (as expected for independent pathway effects); BLISS value > 20 indicates activity greater than additive (synergy); and BLISS value < 0 indicates the combination is less than additive (antagonism).

## RT-qPCR

Calu-3 cells (750,000 cells per well) were plated in six-well plates. The next day, drugs were added to cells. One hour later, cells were infected with SARS-CoV-2 (MOI = 0.3). For ALI cultures, the apical surface was washed with OptiMEM, and cells were placed into fresh medium with drugs added to the basolateral surface. Cells were infected apically with SARS-CoV-2 (MOI = 0.2) for 1 hour and, subsequently, the virus inoculum was removed. The cells were placed into fresh medium daily with the indicated drugs. Total RNA was purified using TRIzol (Invitrogen) followed by the RNA Clean and Concentrate kit (Zymo Research) 48 h after infection for Calu-3 and 72 h after infection for ALI cultures. For cDNA synthesis, reverse transcription was performed with random hexamers and Moloney murine leukaemia virus (M-MLV) reverse transcriptase (Invitrogen). Synthesized RNA was used as a standard (BEI). Gene-specific primers to SARS-CoV-2 (Wuhan v1, NSP14) and SYBR green master mix (Applied Biosystems) were used to amplify viral RNA, and 18S rRNA primers were used to amplify cellular RNA using the QuantStudio 6 Flex RT-PCR system (Applied Biosystems). Relative quantities of viral and cellular RNA were calculated using the standard curve method<sup>3</sup>. Viral RNA was normalized to 18S RNA for each sample (Wuhan V1/18S)<sup>4</sup>.

## Metabolomics

Calu-3 cells treated for 24 h with the indicated drugs were snap-frozen in liquid nitrogen, and thawed cell pellets on ice were homogenized with ice-cold 80% methanol. The samples were vortexed for 10 s to lyse cells and the cell homogenate was spiked with isotopically labelled nucleotide internal standards. The mixture was extracted with ice-cold methanol, vortexed and centrifuged at 18,100g for 5 min at 4

°C. The supernatant was dried under nitrogen at 45 °C and reconstituted in 50% methanol for liquid chromatography–mass spectrometry on an Agilent 1290 Infinity UHPLC/6495B triple quadrupole mass spectrometer. A 12-min linear gradient from 95% B (acetonitrile) to 54% A (10 mM ammonium acetate, pH 9, and 0.1% medronic acid) on an Agilent PEEK HILIC-z 2 × 100 mm, 2.7- $\mu$ m column was used to separate nucleotides. Multiple reaction monitoring was used to quantitate a fragment ion of the parent ion of each nucleotide with standard calibration curves. The raw data were normalized from  $\mu$ M to nmol mg<sup>-1</sup> protein using a BCA assay from input samples.

## Mouse studies

Mouse challenge studies were approved by the University of Maryland School of Medicine IACUC. Eight to ten-week-old female BALB/C mice were randomized and not blinded. Molnupiravir (#HY-135853, MedChemExpress) was resuspended in corn oil (8267, Sigma) and 10% DMSO (2438, Sigma), with dosing twice a day as oral gavage. Brequinar (#HY-108325, MedChemExpress) was resuspended in 10% DMSO and sterile saline, with dosing daily as intraperitoneal injection. Our pharmacokinetic (PK) studies showed an approximately 10 h half-life of Brequinar, and dosing was either started at 12 h before infection or 24 h after infection as indicated.

Mice were anaesthetized by intraperitoneal injection with 50  $\mu$ l of a mix of xylazine (0.38 mg per mouse) and ketamine (1.3 mg per mouse) diluted in PBS. Mice were intranasally inoculated with  $1 \times 10^5$  p.f.u. of the Beta variant of SARS-CoV-2 in 50  $\mu$ l. Challenged mice were weighed on the day of infection and daily for 2 days after infection; there were no significant changes in weights observed. For prophylactic dosing, 2 days after infection, or therapeutic dosing 3 days after infection, five mice were killed from each treatment and control group, lungs were collected to determine viral titre by a plaque assay, and fixed in 4% paraformaldehyde for 24 h before sectioning and staining with haematoxylin and eosin by UMSOM Histology Core<sup>5</sup>. Pathological scoring on blinded haematoxylin and eosin-stained sections was performed for each mouse and analysed for inflammation.

## Statistical analyses

Statistics were performed with GraphPad Prism 9.2 software. Lung titres for the treatment groups were compared to vehicle and statistically analysed using a non-parametric one-way ANOVA with Dunnett's multiple comparison test. For metabolomics, one-way ANOVA with Dunnett's post-hoc test for multiple comparisons was used to determine significance from vehicle control. For RT-qPCR, two-way ANOVA test was used. For the synergy studies, Combenefit software (update 1 June 2016) was used. Statistical methods were not used to predetermine sample size. Blinding and randomization were not used.

## Reporting summary

Further information on research design is available in the [Nature Research Reporting Summary](#) linked to this paper.

## Data availability

All data in this paper are presented in the main text and supplementary text. [Supplementary Information](#) is available for this paper. [Source data](#) are provided with this paper.

## References

1. Lu, R. et al. Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. *Lancet* **395**, 565–574 (2020).
2. Huang, C. et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet* **395**, 497–506 (2020).
3. Huchting, J. Targeting viral genome synthesis as broad-spectrum approach against RNA virus infections. *Antivir. Chem. Chemother.* **28**, 2040206620976786 (2020).
4. Jordheim, L. P., Durantel, D., Zoulim, F. & Dumontet, C. Advances in the development of nucleoside and nucleotide analogues for cancer and viral diseases. *Nat. Rev. Drug Discov.* **12**, 447–464 (2013).
5. Sheahan, T. P. et al. Comparative therapeutic efficacy of remdesivir and combination lopinavir, ritonavir, and interferon beta against MERS-CoV. *Nat. Commun.* **11**, 222 (2020).
6. Agostini, M. L. et al. Coronavirus susceptibility to the antiviral remdesivir (GS-5734) is mediated by the viral polymerase and the proofreading exoribonuclease. *mBio* **9**, e00221-18 (2018).
7. Pruijssers, A. J. et al. Remdesivir inhibits SARS-CoV-2 in human lung cells and chimeric SARS-CoV expressing the SARS-CoV-2 RNA polymerase in mice. *Cell Rep.* **32**, 107940 (2020).
8. Sheahan, T. P. et al. Broad-spectrum antiviral GS-5734 inhibits both epidemic and zoonotic coronaviruses. *Sci. Transl. Med.* **9**, eaal3653 (2017).

9. Dittmar, M. et al. Drug repurposing screens reveal cell-type-specific entry pathways and FDA-approved drugs active against SARS-CoV-2. *Cell Rep.* **35**, 108959 (2021).
10. Janes, J. et al. The ReFRAME library as a comprehensive drug repurposing library and its application to the treatment of cryptosporidiosis. *Proc. Natl Acad. Sci. USA* **115**, 10750–10755 (2018).
11. Bakowski, M. A. et al. Drug repurposing screens identify chemical entities for the development of COVID-19 interventions. *Nat. Commun.* **12**, 3309 (2021).
12. Riva, L. et al. Discovery of SARS-CoV-2 antiviral drugs through large-scale compound repurposing. *Nature* **586**, 113–119 (2020).
13. Yates, M. K. & Seley-Radtke, K. L. The evolution of antiviral nucleoside analogues: a review for chemists and non-chemists. Part II: complex modifications to the nucleoside scaffold. *Antiviral Res.* **162**, 5–21 (2019).
14. Sheahan, T. P. et al. An orally bioavailable broad-spectrum antiviral inhibits SARS-CoV-2 in human airway epithelial cell cultures and multiple coronaviruses in mice. *Sci. Transl. Med.* **12**, eabb5883 (2020).
15. Jordan, P. C. et al. Initiation, extension, and termination of RNA synthesis by a paramyxovirus polymerase. *PLoS Pathog.* **14**, e1006889 (2018).
16. Tchesnokov, E. P., Feng, J. Y., Porter, D. P. & Gotte, M. Mechanism of inhibition of Ebola virus RNA-dependent RNA polymerase by remdesivir. *Viruses* **11**, 326 (2019).
17. Agostini, M. L. et al. Small-molecule antiviral  $\beta$ -d- $N^4$ -hydroxycytidine inhibits a proofreading-intact coronavirus with a high genetic barrier to resistance. *J. Virol.* **93**, e01348-19 (2019).
18. Urakova, N. et al.  $\beta$ -d- $N^4$ -hydroxycytidine is a potent anti-alphavirus compound that induces a high level of mutations in the viral genome. *J. Virol.* **92**, e01965-17 (2018).
19. Ma, J. & Motsinger-Reif, A. Current methods for quantifying drug synergism. *Proteom. Bioinform.* **1**, 43–48 (2019).
20. Ariav, Y., Ch'ng, J. H., Christofk, H. R., Ron-Harel, N. & Erez, A. Targeting nucleotide metabolism as the nexus of viral infections, cancer, and the immune response. *Sci. Adv.* **7**, eabg6165 (2021).

21. Wahl, A. et al. SARS-CoV-2 infection is effectively treated and prevented by EIDD-2801. *Nature* **591**, 451–457 (2021).
22. Cox, R. M., Wolf, J. D. & Plemper, R. K. Therapeutically administered ribonucleoside analogue MK-4482/EIDD-2801 blocks SARS-CoV-2 transmission in ferrets. *Nat. Microbiol.* **6**, 11–18 (2021).
23. Hou, Y. J. et al. SARS-CoV-2 reverse genetics reveals a variable infection gradient in the respiratory tract. *Cell* **182**, 429–446.e14 (2020).
24. Pizzorno, A. et al. Characterization and treatment of SARS-CoV-2 in nasal and bronchial human airway epithelia. *Cell Rep. Med.* **1**, 100059 (2020).
25. Tyson, J. R. et al. Improvements to the ARTIC multiplex PCR method for SARS-CoV-2 genome sequencing using nanopore. Preprint at bioRxiv  
<https://doi.org/10.1101/2020.09.04.283077> (2020).
26. Li, H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* **34**, 3094–3100 (2018).
27. Robinson, J. T. et al. Integrative genomics viewer. *Nat. Biotechnol.* **29**, 24–26 (2011).

## Acknowledgements

We thank members of the Cherry, Lynch, Frieman and Penn Center for Precision Medicine for advice and discussion; R. Perera for viral preparations; D. Nguyen for his data analysis and all members of the High-Throughput Screening Core at University of Pennsylvania for reagents and technical support; the Penn Metabolomics Core in the Cardiovascular Institute for liquid chromatography–mass spectrometry quantitation of nucleotides; S. Weiss and Y. Li for sharing SARS-related coronavirus 2, isolate USA-WA1/2020; A. Pekosz for SARS-CoV-2 variants (Alpha, Beta, Gamma and Delta); X. Hu, R. Eastman, M. Hall and members of NCATS Compound Management for assembly and preparation of the NCATS compound collections; and CALIBR for providing the ReFrame library and validation plates. This work was supported by grants from the National Institutes of Health to S. Cherry (R01AI074951, R01AI122749, 1R21AI151882 and R01AI140539), as well as funding from the Penn Center for Precision Medicine, Mercatus and the Bill and Melinda Gates Foundation (INV-018479). S. Cherry is a recipient of the Burroughs Wellcome Investigators in the Pathogenesis of Infectious Disease Award, the Deans Innovation Fund as well as Linda and Laddy Montague. Work at the NCATS was funded by the Intramural Research Program of the National Center for Advancing Translational Sciences, National Institutes of Health (ZIA Project #TR000414-01).

M.B.F., R.M.J., H.L.H. and S.W. are supported by funding from Bill and Melinda Gates Foundation (INV-016638), NIH R21AI158134, NIH R21AI153480 and HHS/BARDA ASPR-20-01495.

## Author information

### Author notes

1. These authors contributed equally: Robert M. Johnson, Kasirajan Ayyanathan, Jesse Miller

### Affiliations

1. Department of Biochemistry and Biophysics, University of Pennsylvania, Philadelphia, PA, USA

David C. Schultz, Kanupriya Whig, Brinda Kamalia & Sara Cherry

2. Department of Microbiology and Immunology, Center for Pathogen Research, University of Maryland School of Medicine, Baltimore, MD, USA

Robert M. Johnson, Stuart Weston, Holly L. Hammond, Carly Dillen, Jeremy Ardanuy, Louis Taylor & Matthew B. Frieman

3. Department of Pathology and Laboratory Medicine, University of Pennsylvania, Philadelphia, PA, USA

Kasirajan Ayyanathan, Jesse Miller, Mark Dittmar, Jae Seung Lee, Minghua Li & Sara Cherry

4. National Center for Advancing Translational Sciences, National Institutes of Health, Rockville, MD, USA

Emily Lee & Marc Ferrer

5. Metabolomics Core, Penn Cardiovascular Institute, University of Pennsylvania, Philadelphia, PA, USA

Clarissa Shoffler & Christopher Petucci

6. Epithelix, Geneva, Switzerland

Samuel Constant

7. Department of Microbiology, University of Pennsylvania, Philadelphia, PA, USA

Christoph A. Thaiss & Sara Cherry

## Contributions

D.C.S., R.M.J., E.L., S. Constant, M.F., M.B.F. and S. Cherry. conducted research design. D.C.S., R.M.J., K.A., J.M., K.W., B.K., M.D., S.W., H.L.H., C.D., J.A., L.T., J.S.L., M.L., E.L., C.S., C.P., S. Constant, M.F., M.B.F. and S. Cherry. performed the experiments. D.C.S., R.M.J., K.A., J.M., K.W., B.K., M.D., S.W., H.L.H., C.D., J.S.L., M.L., E.L., C.P., S. Constant, M.F., M.B.F. and S. Cherry conducted the data analysis. D.C.S., E.L., S. Constant, M.F., C.A.T., M.B.F. and S. Cherry wrote the manuscript. D.C.S., S. Constant, M.F., M.B.F. and S. Cherry. supervised the study.

## Corresponding authors

Correspondence to [David C. Schultz](#), [Matthew B. Frieman](#) or [Sara Cherry](#).

## Ethics declarations

## Competing interests

The authors declare no competing interests.

## Peer review

### Peer review information

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

## Additional information

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## Extended data figures and tables

[Extended Data Fig. 1 High-throughput screening in Calu-3 cells to identify antivirals against SARS-CoV-2.](#)

**a–c**, High throughput screening of ~18,000 drugs including the **(a)** UPENN library, **(b)** NCATS library and **(c)** ReFrame library in Calu-3 cells infected with SARS-CoV-2. Percent of Control (POC) for % infection is plotted versus rank of the drug across the primary screens. Yellow circles, DMSO controls; Green circles, remdesivir controls; Blue Circles, sample wells. Solid horizontal line represents 100% Infection of DMSO control wells. **d**, Huh7.5 or Vero (CCL81) cells were treated with the indicated nucleosides in dose response showing infection (blue) and toxicity (green). Data are presented as mean ± SD (n = 3 independent biological replicates). **e**, The indicated cell line was treated with increasing doses of 6-mercaptopurine and dose responses for infection (blue) and toxicity (green) are shown. Data are presented as mean ± SD (n = 3 independent biological replicates). **f**, p-values from Fig. [1d](#).

### [Extended Data Fig. 2 Structures of antiviral nucleoside analogs shown.](#)

Nucleoside-related analog structures shown.

### [Extended Data Fig. 3 Combination of remdesivir and molnupiravir is additive.](#)

**a**, The indicated cell line was treated with increasing doses of EIDD-1931 and dose responses for infection (blue) and toxicity (green) are shown. Data are presented as mean ± SD (n = 3 independent biological replicates). **b**, anti-SARS2 activity of remdesivir and molnupiravir as single agents. Data are presented as mean ± SD. **c**, anti-SARS2 activity of remdesivir in combination with a single concentration of molnupiravir or molnupiravir in combination with a single concentration of remdesivir. **d**, Two-dimensional representation of dose response interaction matrix for percent of control of infection mean ± SD, n = 9 independent biological replicates.

### [Extended Data Fig. 4 Inhibitors of host nucleoside metabolism are antiviral.](#)

**a**, Calu-3 cells were treated with the indicated IMPDH inhibitors in dose response showing infection (blue) and toxicity (green). Data are presented as mean ± SD (n = 4 independent biological replicates). **b**, A549-Ace2 cells were treated with the indicated inhibitors in dose response showing infection (blue) and toxicity (green). Data are presented as mean ± SD (n = 4 independent biological replicates). **c**, Table of the EC50, CC50 and SI for the indicated cell line and drug (uM).

### [Extended Data Fig. 5 Combination of molnupiravir or remdesivir with DHODH inhibitors are synergistically antiviral in vitro.](#)

**a**, Anti-SARS2 activity of molnupiravir, remdesivir, and Brequinar, as single agents. Data are presented as mean  $\pm$  SD. **b**, Anti-SARS2 activity of molnupiravir or remdesivir in combination with a single concentration of Brequinar. **c**, Two-dimensional representation of dose response interaction matrix for percent of control of infection. Data are presented as mean  $\pm$  SD (n = 6 independent biological replicates). **d**, Anti-SARS2 activity of molnupiravir, remdesivir, and BAY-2402234, as single agents. Data are presented as mean  $\pm$  SD. **e**, Anti-SARS2 activity of molnupiravir or remdesivir in combination with a single concentration of BAY-2402234. **f**, Two-dimensional representation of dose response interaction matrix for percent of control of infection. Data are presented as mean  $\pm$  SD (n = 6 for EIDD-2801, n = 8 for remdesivir for independent biological replicates).

**Extended Data Fig. 6 Combination of molnupiravir or remdesivir with Pyrazofurin but not AVN944 is synergistically antiviral in vitro.**

**a**, BLISS analysis in Calu3 cells with molnupiravir (EIDD-2801) or remdesivir in combination with pyrazofurin. Data are presented as mean values of Excess over BLISS. (n = 4 independent biological replicates). The statistical significance was determined by a one-sample t-test (\* p < 5x10<sup>-2</sup>; \*\*p < 10<sup>-3</sup>). **b**, Anti-SARS2 activity of molnupiravir, remdesivir, and AVN944 as single agents. Data are presented as mean  $\pm$  SD. **c**, Anti-SARS2 activity of molnupiravir or remdesivir in combination with a single concentration of AVN944. **d**, Two-dimensional representation of dose response interaction matrix for percent of control of infection. Data are presented as mean  $\pm$  SD (n = 10 for EIDD-2801, n = 6 for remdesivir for independent biological replicates).

**Extended Data Fig. 7 Combination of molnupiravir with DHODH inhibitors is synergistic against SARS-CoV-2 variants.**

**a**, Table of EC50, CC50 and SI for the indicated variant and the indicated drugs (uM) in Calu-3 cells. **b**, BLISS analysis in Calu3 cells with molnupiravir (EIDD-2801) in combination with DHODH inhibitors Brequinar or BAY-2402234 infected with the indicated variant. Data are presented as mean values of Excess over BLISS. The following number of independent replicates were performed: SARS-CoV2 Alpha; n = 8 for BAY2402234, n = 6 Brequinar; SARS-CoV2 Beta; n = 7, SARS-CoV2 Gamma; n = 6 BAY2402234, n = 8; SARS-CoV2; Delta: n = 8. The statistical significance was determined by a one-sample t-test (\* p < 5x10<sup>-2</sup>; \*\*p < 10<sup>-3</sup>, \*\*\*p < 10<sup>-4</sup>).

**Extended Data Fig. 8 Nucleoside-related drugs are well-tolerated in air-liquid interface cultures.**

**a**, Calu-3 cells were treated with the indicated concentrations of Molnupiravir (EIDD-2801) and infected with SARS-CoV-2. 48hpi viral replication was quantified by RT-qPCR and expression (viral RNA/18S) was normalized to vehicle treated cells. Data are presented as mean±SEM for n = 2-3 biological replicates. **b**, P-values for Fig. 3d. **c**, Nasal air-liquid interface cells were treated daily on the basolateral side with the indicated concentrations of drug, and at 72 h TEER, Cilia beating frequency (CBF) and toxicity (LDH) were measured. **d**, P-values for Fig. 4.

### **Extended Data Fig. 9 Combination of molnupiravir and Brequinar reduces SARS-CoV-2 infection and inflammation in vivo during therapeutic dosing.**

Wild type Balb/C mice were treated with Brequinar (IP) and/or molnupiravir (EIDD-2801) (PO) daily at the indicated concentrations starting 24 h after infection. Mice (N = 5 mice per group) were intranasally inoculated with  $1 \times 10^5$  PFU/mouse of SARS-CoV-2/B.1.351. **a**, 3 dpi, lungs analyzed for viral titer by plaque assay, Mean±SD shown. \* p < 0.05, \*\* p < 0.01, non-parametric one-way ANOVA with Dunnett's multiple comparison test. (p-values: B1.135 only vs EIDD-2801 (25 mg/kg) = 0.045, vs EIDD-2801 (50 mg/kg) = 0.016, vs EIDD-2801 (25 mg/kg)+ Breq (50 mg/kg) = 0.0017. **b, c**, 3 dpi, lungs fixed in 4% PFA for H&E staining (Scale bar = 200  $\mu$ M) and quantified for interstitial inflammation. Mean±SD shown. \*\* p < 0.01, non-parametric one-way ANOVA with Dunnett's multiple comparison test. (p-value = 0.0015).

#### [Source data](#)

### **Extended Data Fig. 10 Combination of molnupiravir or remdesivir with the protease inhibitor PF-07321332 is additive against SARS-CoV-2 (Beta, Delta).**

BLISS analysis in Calu-3 cells with molnupiravir (EIDD-2801) or remdesivir in combination with PF-07321332 infected with the indicated SARS-CoV-2 variant. Data are presented as mean Excess over BLISS for n = 2 independent biological replicates (SARS-CoV-2 Beta); n = 6 independent biological replicates (SARS-CoV-2 Delta). The statistical significance was determined by a one-sample t-test (\* p <  $5 \times 10^{-2}$ ; \*\*p <  $10^{-3}$ ).

## **Supplementary information**

### [Supplementary Information](#)

This file contains Supplementary Table 1.

## Reporting Summary

## Source data

### Source Data Fig. 4

### Source Data Extended Data Fig. 9

## Rights and permissions

### Reprints and Permissions

## About this article

### Cite this article

Schultz, D.C., Johnson, R.M., Ayyanathan, K. *et al.* Pyrimidine inhibitors synergize with nucleoside analogues to block SARS-CoV-2. *Nature* **604**, 134–140 (2022).  
<https://doi.org/10.1038/s41586-022-04482-x>

- Received: 22 October 2021
- Accepted: 26 January 2022
- Published: 07 February 2022
- Issue Date: 07 April 2022
- DOI: <https://doi.org/10.1038/s41586-022-04482-x>

## Share this article

Anyone you share the following link with will be able to read this content:

[Get shareable link](#)

Sorry, a shareable link is not currently available for this article.

[Copy to clipboard](#)

## Further reading

- [A promising strategy against SARS-CoV-2: pyrimidine inhibitors synergize with nucleoside analogues](#)

- Long Min
- Qiu Sun

*Signal Transduction and Targeted Therapy* (2022)

- [Inhibition of the hexamerization of SARS-CoV-2 endoribonuclease and modeling of RNA structures bound to the hexamer](#)

- Duy Phuoc Tran
- Yuta Taira
- Akio Kitao

*Scientific Reports* (2022)

---

This article was downloaded by **calibre** from <https://www.nature.com/articles/s41586-022-04482-x>

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

[Download PDF](#)

- Article
- [Published: 15 February 2022](#)

# Germinal centre-driven maturation of B cell response to mRNA vaccination

- [Wooseob Kim](#) [ORCID: orcid.org/0000-0002-9199-1000](#)<sup>1</sup><sup>nal</sup>,
- [Julian Q. Zhou](#) [ORCID: orcid.org/0000-0001-9602-2092](#)<sup>1</sup><sup>nal</sup>,
- [Stephen C. Horvath](#)<sup>1</sup>,
- [Aaron J. Schmitz](#) [ORCID: orcid.org/0000-0002-8077-6751](#)<sup>1</sup>,
- [Alexandria J. Sturtz](#)<sup>1</sup>,
- [Tingting Lei](#)<sup>1</sup>,
- [Zhuoming Liu](#) [ORCID: orcid.org/0000-0001-8198-0976](#)<sup>2</sup>,
- [Elizaveta Kalaidina](#)<sup>3</sup>,
- [Mahima Thapa](#)<sup>1</sup>,
- [Wafaa B. Alsoussi](#)<sup>1</sup>,
- [Alem Haile](#)<sup>4</sup>,
- [Michael K. Klebert](#)<sup>4</sup>,
- [Teresa Suessen](#)<sup>5</sup>,
- [Luis Parra-Rodriguez](#)<sup>6</sup>,
- [Philip A. Mudd](#) [ORCID: orcid.org/0000-0002-3860-5473](#)<sup>7,8</sup>,
- [Sean P. J. Whelan](#) [ORCID: orcid.org/0000-0003-1564-8590](#)<sup>2</sup>,
- [William D. Middleton](#)<sup>5</sup>,
- [Sharlene A. Teeffey](#)<sup>5</sup>,
- [Iskra Pusic](#)<sup>9</sup>,
- [Jane A. O'Halloran](#) [ORCID: orcid.org/0000-0001-8265-9471](#)<sup>6</sup>,
- [Rachel M. Presti](#) [ORCID: orcid.org/0000-0002-5469-0727](#)<sup>6,8</sup>,
- [Jackson S. Turner](#)<sup>1</sup> &
- [Ali H. Ellebedy](#) [ORCID: orcid.org/0000-0002-6129-2532](#)<sup>1,8,10</sup>

*Nature* volume 604, pages 141–145 (2022)

- 22k Accesses
- 995 Altmetric

- [Metrics details](#)

## Subjects

- [Antibodies](#)
- [Germinal centres](#)
- [Plasma cells](#)
- [Somatic hypermutation](#)

## Abstract

Germinal centres (GC) are lymphoid structures in which B cells acquire affinity-enhancing somatic hypermutations (SHM), with surviving clones differentiating into memory B cells (MBCs) and long-lived bone marrow plasma cells<sup>[1,2,3,4,5](#)</sup> (BMPCs). SARS-CoV-2 mRNA vaccination induces a persistent GC response that lasts for at least six months in humans<sup>[6,7,8](#)</sup>. The fate of responding GC B cells as well as the functional consequences of such persistence remain unknown. Here, we detected SARS-CoV-2 spike protein-specific MBCs in 42 individuals who had received two doses of the SARS-CoV-2 mRNA vaccine BNT162b2 six month earlier. Spike-specific IgG-secreting BMPCs were detected in 9 out of 11 participants. Using a combined approach of sequencing the B cell receptors of responding blood plasmablasts and MBCs, lymph node GC B cells and plasma cells and BMPCs from eight individuals and expression of the corresponding monoclonal antibodies, we tracked the evolution of 1,540 spike-specific B cell clones. On average, early blood spike-specific plasmablasts exhibited the lowest SHM frequencies. By contrast, SHM frequencies of spike-specific GC B cells increased by 3.5-fold within six months after vaccination. Spike-specific MBCs and BMPCs accumulated high levels of SHM, which corresponded with enhanced anti-spike antibody avidity in blood and enhanced affinity as well as neutralization capacity of BMPC-derived monoclonal antibodies. We report how the notable persistence of the GC reaction induced by SARS-CoV-2 mRNA vaccination in humans culminates in affinity-matured long-term antibody responses that potently neutralize the virus.

## Main

Vaccination of humans with the Pfizer-BioNTech SARS-CoV-2 mRNA vaccine BNT162b2 induces a robust but transient circulating plasmablast response and a persistent germinal centre (GC) reaction in the draining lymph nodes<sup>[6](#)</sup>. Whether these persistent GC responses lead to the generation of affinity-matured memory B cells (MBCs) and long-lived bone marrow-resident plasma cells (BMPCs) remains unclear. To address this question, we analysed long-term B cell responses in the participants

enrolled in our previously described observational study of 43 healthy participants (13 with a history of SARS-CoV-2 infection) who received two doses of BNT162b2<sup>6,7</sup> (Extended Data Tables 1). Long-term blood samples ( $n = 42$ ) and fine needle aspirates (FNAs) of the draining axillary lymph nodes ( $n = 15$ ) were collected 29 weeks after vaccination (Fig. 1a). Bone marrow aspirates were collected 29 ( $n = 11$ ) and 40 weeks ( $n = 2$ ) after vaccination, with the 40-week time point used only for B cell receptor (BCR) repertoire profiling (Fig. 1a). None of the participants who contributed FNA or bone marrow specimens had a history of SARS-CoV-2 infection.

**Fig. 1: Persistence of humoral immune responses to SARS-CoV-2 mRNA vaccination.**

---

 **figure 1**

---

**a**, Forty-three participants (13 with a history of SARS-CoV-2 infection) were enrolled, followed by vaccination. Blood ( $n = 42$ ) was collected before and at indicated time points after vaccination. For 15 participants without infection history, aspirates of draining axillary lymph nodes were collected at indicated time points after

vaccination. For 11 participants without infection history, aspirates of bone marrow were collected at 29 and 40 weeks after vaccination. **b**, Representative flow cytometry plots of GC B cells ( $CD19^+CD3^-IgD^{low}BCL6^+CD38^{int}$ ) and S-binding GC B cells in lymph nodes 29 weeks after vaccination. The percentage of cells in the bound region is indicated. **c**, Kinetics of total (left) and S-binding GC B cells (right) as gated in **b**. **d**, Representative ELISpot wells coated with the indicated antigens, bovine serum albumin or anti-immunoglobulin and developed in blue (IgG) and red (IgA) after plating the indicated numbers of BMPCs. **e**, Frequencies of IgG-secreting BMPCs specific for the indicated antigens 29 weeks after vaccination. Symbols at each time point represent one sample in **c** ( $n = 15$ ) and **e** ( $n = 11$ ). **f**, Plasma anti-S IgG titres measured by ELISA in participants without (red,  $n = 29$ ) and with (black,  $n = 9$ ) infection history. Horizontal lines and numbers indicate geometric means. Results are from one experiment performed in duplicate. **g**, Representative flow cytometry plot of S-binding MBCs ( $CD20^+CD38^-IgD^{low}CD19^+CD3^-$ ) in blood 29 weeks after vaccination. **h**, Frequencies of S-specific MBCs in participants without (red,  $n = 29$ ) and with (black,  $n = 13$ ) infection history as gated in **g**. Horizontal lines indicate medians in **e** and **h**. LoD, limit of detection.

GC B cells were detected in FNAs from all 15 participants (Fig. [1b,c](#), left, Extended Data Fig. [1a](#), Extended Data Table [2](#)). All 14 participants with FNAs collected prior to week 29 generated spike (S)-binding GC B cell responses of varying magnitudes (Fig [1b,c](#), right, Extended Data Table [2](#)). Notably, S-binding GC B cells were detected in FNAs from 10 out of 15 participants at week 29 (Fig. [1b,c](#), right, Extended Data Table [2](#)), demonstrating that two-thirds of the sampled participants maintained an antigen-specific GC B cell response for at least six months after vaccination. S-binding lymph node plasma cells (LNPCs) were also detected in FNAs from all 15 participants and exhibited similar dynamics to S-binding GC B cells, albeit at lower frequencies within the total B cell population (Extended Data Fig. [1a,b](#), Extended Data Table [2](#)). None of the FNAs demonstrated significant contamination with peripheral blood, based on the nearly complete absence of myeloid cells (Extended Data Table [2](#)).

Frequencies of BMPCs secreting IgG or IgA antibodies against either the 2019–2020 inactivated influenza virus vaccine, the tetanus–diphtheria vaccine or S protein were assessed in bone marrow aspirates collected 29 weeks after vaccination by enzyme-linked immunosorbent spot assay (ELISpot) (Fig. [1d,e](#), Extended Data Fig. [1c](#)). Influenza and tetanus–diphtheria vaccine-specific IgG-secreting BMPCs were detectable (median frequencies of 1.4% and 0.15%, respectively) in all 11 participants (Fig. [1e](#)). S-binding IgG-secreting BMPCs were detected in 9 out of 11 participants (median frequency of 0.06%). IgA-secreting BMPCs specific to influenza vaccine were detected in 10 out of 11 participants, but IgA-secreting BMPCs directed against the tetanus–diphtheria vaccine and the S protein were largely below the limit of

detection (Extended Data Fig. 1c). All participants had detectable plasma anti-S IgG antibodies and circulating S-binding MBCs at the 29-week time point (Fig. 1f-h). Anti-S IgG titres at 29 weeks were higher than titres observed in a cohort of unvaccinated people who had recovered from SARS-CoV-2 measured 29 weeks after infection<sup>9,10,11</sup> (Extended Data Fig. 1d). Vaccinated participants with a history of SARS-CoV-2 infection had significantly higher titres of anti-S IgG antibodies at 5 and 29 weeks compared with their naive counterparts<sup>9,11,12</sup> (Fig. 1f). Similar trends were observed for plasma anti-S IgM and IgA antibodies (Extended Data Fig. 1e). S-binding MBCs were detected in all participants, with a median frequency of 0.23% of total circulating B cells (Fig. 1g,h, Extended Data Fig. 1f).

To track S-specific B cell evolution and clonal distribution within blood, lymph node and bone marrow, we performed single-cell RNA sequencing (scRNA-seq) and concurrent BCR sequencing of immune cells from eight participants who contributed specimens from the three compartments. We first sorted plasmablasts from samples collected at their peak frequencies, one week after the second immunization<sup>6</sup> (Fig. 2a, top, Extended Data Fig. 2a). We then investigated the dynamics of the immune response in draining axillary lymph nodes. Single-cell transcriptional analysis of lymph nodes revealed distinct immune cell populations, as previously described<sup>13,14,15,16</sup> (Fig. 2a, bottom left, Extended Data Fig. 2b,c, Extended Data Table 3). To further distinguish distinct B cell subsets in the lymph node, we performed unbiased secondary clustering of the B cell populations from the total cellular analysis (Fig. 2a, bottom right, Extended Data Fig. 2d,e, Extended Data Table 3). Around 40% and 7.9% of the B cells in the lymph node had GC B cell and LNPC transcriptomic profiles, respectively.

**Fig. 2: Identification of SARS-CoV-2 S-binding B cell clones in draining axillary lymph nodes.**



**a**, Uniform manifold approximation and projection (UMAP) showing scRNA-seq transcriptional clusters of total cells (left) and of B cells (right) from plasmablasts (PBs) sorted from PBMC (top) and from FNA of lymph nodes (bottom). Each dot represents a cell, coloured by phenotype as defined by transcriptomic profile. Total numbers of cells are at the top right corner. FDC, follicular dendritic cell; GC, GC B cell; Mo, monocyte; NK, natural killer cell; PB, plasmablast; pDC, plasmacytoid dendritic cell. **b**, Positive binding of recombinant monoclonal antibodies derived from GC B cells (blue) or LNPCs (green) to SARS-CoV-2 S measured by ELISA. Results are from one experiment performed in duplicate.

We next generated recombinant monoclonal antibodies from expanded clones detected in FNA samples 7 and 15 weeks after vaccination, representing early and late time points. For two of the eight participants from whom the late point was unavailable due to insufficient specimens, we analysed two separate early time points: weeks five and seven for participant 02a, and weeks four and seven for participant 04. A total of 2,099 recombinant monoclonal antibodies were generated, of which 1,503 (71.6%) bound SARS-CoV-2 S in an enzyme-linked immunosorbent assay (ELISA) (Fig. 2b, Extended Data Table 4). In subsequent analyses, we included 37 previously identified S-binding monoclonal antibodies generated from GC B cells at week 4 from participants 07, 20, and 22<sup>6</sup>. Clonal relationships were computationally inferred using heavy chains from scRNA-seq BCR libraries (Extended Data Table 5), bulk-seq BCR libraries for GC B cells, LNPCs (Extended Data Fig. 2g) and BMPCs (Extended Data Table 5), as well as previously published bulk-seq BCR libraries of sorted plasmablasts and GC B cells<sup>6</sup>, and magnetically enriched IgD<sup>low</sup> activated B cells or MBCs from PBMC<sup>17</sup>. B cell clones with experimentally validated S-binding B cells were designated S-binding clones (Extended Data Fig. 2f) and accounted for 43.1% and 64.4%, respectively of the single-cell profiled GC B cells and LNPCs (Extended Data Fig. 2h, Extended Data Table 3). B cells that were clonally related to S-binding B cells were also found in the plasmablast compartment in blood (6.7%) and the MBC compartment in lymph nodes (0.3%) (Extended Data Fig. 2h, Extended Data Table 3).

## B cell maturation in the germinal centre

We analysed the proportion of S-binding GC B cells clonally related to week 4 circulating plasmablasts. The frequencies of plasmablast-related, S-binding GC B cells varied broadly among participants, ranging from 12.7% to 82.5% (Fig. 3a). Consistent with our flow cytometry results (Fig. 1c), GC B cells from long-lasting S-binding clones were observed for at least 29 weeks after vaccination (Extended Data Fig. 3a). In addition, we detected the presence of clonally related MBCs in blood at 29 weeks after vaccination (Extended Data Fig. 3b). S-binding GC B cells accumulated significantly higher levels of SHM compared to clonally related plasmablasts, and this difference increased over time (Fig. 3b). We observed a 3.5-fold increase in SHM

frequency among all S-binding GC B cells between weeks 4 and 29 (Fig. 3c, Extended Data Fig. 3c). S-binding MBCs detected at 29 weeks after vaccination, however, had slightly lower SHM frequencies than their clonally related GC B cell counterparts (Extended Data Fig. 3d). The relative proportion of S-binding GC B cells expressing BCR of IgA isotype increased in the lymph node over time (Extended Data Fig. 3e). Clonal analysis revealed a high degree of overlap between S-binding GC and LNPC compartments (Fig. 3d). Furthermore, SHM frequencies of both S-binding LNPsCs and GC B cells increased over time at a very similar rate with small differences (Fig. 3e) in contrast to those between S-binding plasmablast and GC B cells (Fig. 3b).

**Fig. 3: Maturation of SARS-CoV-2 S-binding B cells in the lymph node.**

 figure 3

**a**, Circos diagrams showing clonal overlap between S-binding plasmablasts and GC B cells at indicated time points. Purple and grey chords correspond to clones spanning both compartments and clones spanning only the GC compartment, respectively. Percentages are of GC B cell clones related to plasmablasts at each time point. **b**, Nucleotide mutation frequency in the immunoglobulin heavy chain variable gene (*IGHV*) region for clonally related week-4 plasmablasts and GC B cells at weeks 4 ( $n = 81$ ), 5 ( $n = 52$ ), 7 ( $n = 289$ ), 15 ( $n = 162$ ) and 29 ( $n = 47$ ). **c**, *IGHV* nucleotide mutation frequency of S-binding GC B cells at weeks 4 ( $n = 1,701$ ), 5 ( $n = 21,543$ ), 7 ( $n = 62,927$ ), 15 ( $n = 49,837$ ) and 29 ( $n = 3,314$ ). Horizontal lines and numbers represent medians.  $P$  values were determined by Kruskal–Wallis test followed by Dunn’s multiple comparison test. **d**, Circos diagrams showing clonal overlap (purple) between S-binding GC B cells and LNPsCs over combined time points. Percentages are of GC B cell clones overlapping with LNPsCs or vice versa. Arc length corresponds to the number of BCR sequences and chord width corresponds to clone size in **a** and **d**. **e**, *IGHV* nucleotide mutation frequency of clonally related GC B cells and LNPsCs at weeks 4 ( $n = 48$ ), 5 ( $n = 224$ ), 7 ( $n = 877$ ), 15 ( $n = 449$ ) and 29 ( $n = 76$ ). Each dot represents the median SHM frequency of a clone within the indicated compartment, and medians are presented on the top of each dataset in **b** and **e**.  $P$  values were determined by paired two-sided Mann–Whitney test and corrected for multiple testing using the Benjamini–Hochberg method in **b** and **e**. \*\*\* $P < 0.0001$ .

## Affinity maturation of antibody response

To determine whether the increase in SHM frequencies of S-specific GC B cells and LNPCs over time is reflected in increased circulating anti-S antibody binding affinity, we measured the avidity of plasma anti-S IgG. In participants without a history of SARS-CoV-2 infection, anti-S IgG avidity increased at 29 weeks compared with the 5-week time point. Of note, participants with a history of SARS-CoV-2 infection had similar plasma anti-S IgG avidity at 5 and 29 weeks after vaccination (Fig. 4a). Consistently, SHM frequencies of S-binding LNPCs increased over time (Fig. 4b). S-binding BMPCs from 29 and 40 weeks after vaccination exhibited a degree of SHM that was similar to that of LNPCs from 15 and 29 weeks after vaccination (Fig. 4b) and significantly higher than any other S-binding B cell population except for MBCs (Extended Data Fig. 4a). To understand the evolutionary trajectory of vaccine-induced B cell lineages, we analysed S-specific clones using a phylogenetic model tailored for BCR repertoires<sup>18</sup>. Consistent with their SHM frequencies (Fig. 4b), plasmablasts tended to locate closer to the germline on the phylogenetic trees, whereas LNPCs and BMPCs tended to be evolutionarily more distant (Fig. 4c, Extended Data Fig. 4b). In contrast to plasmablasts, which clustered to a separate branch of their own, BMPCs and LNPCs co-located on shared branches, suggesting a closer evolutionary relationship between BMPCs and LNPCs (Fig. 4c). Together, these results support a model in which S-specific BMPCs are the products of affinity-matured, GC-derived LNPCs.

**Fig. 4: Evolution of B cell clones induced by SARS-CoV-2 vaccination.**

---

 **figure 4**

**a**, Avidity indices of plasma anti-S IgG between the indicated time points in participants without (red,  $n = 29$ ) and with (black,  $n = 9$ ) a history of SARS-CoV-2 infection. **b**, *IGHV* nucleotide mutation frequency of S-binding plasmablasts ( $n = 2,735$ ), LNPCs at weeks 4 ( $n = 552$ ), 5 ( $n = 11,253$ ), 7 ( $n = 45,436$ ), 15 ( $n = 24,538$ ) and 29 ( $n = 571$ ), and BMPCs ( $n = 47$ ). Horizontal lines and numbers represent median values.  $P$  values were determined by Kruskal–Wallis test followed by Dunn's multiple comparison test. **c**, Representative phylogenetic trees showing inferred evolutionary relationships between plasmablasts, LNPCs and BMPCs. Horizontal branch length represents the expected number of substitutions per codon in V-region genes, corresponding to the scale bar. Clone IDs are displayed near the root of the trees.

Asterisks denote neutralizing monoclonal antibodies. **d**, Neutralizing activity of clonally related plasmablast- and BMPC-derived monoclonal antibodies ( $n = 8$ ) against SARS-CoV-2 D614G strain. Dotted line indicates detection limit. Results are from one experiment with duplicates in **a** and **d**. **e**, Equilibrium dissociation constant ( $K_D$ ) of neutralizing clone-derived Fabs ( $n = 8$ ) interacting with immobilized S protein measured by BLI. Symbols indicate  $K_D$  values of clonally related, plasmablast (red)- and BMPC (black)-derived Fabs. Results are from at least two replicates in **e**.  $P$  values were determined by two-tailed Wilcoxon matched-pairs signed rank test in **a**, **d** and **e**. NS,  $P > 0.9999$ , \*\*\* $P < 0.0001$ .

We next expressed monoclonal antibodies derived from clonally related plasmablasts and BMPCs and their corresponding monomeric antigen-binding fragments (Fabs) (Extended Data Table 6). We then examined binding affinity and in vitro neutralization capacity using biolayer interferometry (BLI) and high-throughput GFP-reduction neutralization test<sup>19</sup>, respectively. BMPC-derived Fabs exhibited significantly higher binding affinity against S protein compared with plasmablast-derived Fabs (Extended Data Fig. 4c, d). Of the 21 S-specific clones we detected among BMPCs, 7 potently neutralized the SARS-CoV-2 D614G strain (Extended Data Fig. 4e). These BMPC-derived monoclonal antibodies showed higher neutralizing potency than their clonally related, plasmablast-derived counterparts (Fig. 4d), consistent with the significantly increased binding affinity of the BMPC-derived Fabs to S protein (Fig. 4e). Overall, the increased frequency of SHM observed over time and the correlated functional improvements in neutralization suggest that the GC reactions induced by SARS-CoV-2 mRNA vaccination facilitate the development of affinity-matured BMPCs.

## Discussion

This study evaluated whether the persistent GC response induced by SARS-CoV-2 mRNA-based vaccines in humans<sup>6</sup> results in the generation of affinity-matured MBCs and BMPCs<sup>1,3,13,20,21</sup>. The two-dose series of BNT162b2 induced a robust S-binding GC B cell response that lasted for at least 29 weeks after vaccination. The results of such persistent GC reactions were evident in the form of circulating S-binding MBCs in all participants and S-specific BMPCs 29 weeks after vaccination in all but two of the sampled participants. It is likely that S-specific BMPCs in those two participants are present but below the assay detection limit. Longitudinal tracking of more than 1,500 vaccine-induced B cell clones revealed the gradual accumulation of SHM and isotype switching to IgA within the GC B cell compartment. We also show that GC B cells differentiate into affinity-matured LNPs within the lymph node, with some of these cells potentially migrating to the bone marrow where they establish long-term residence. The enhanced maturity of the secreted antibodies was reflected in the significantly increased avidity of circulating anti-S IgG antibodies over time. It is also

evident from increased affinity of BMPC-derived monoclonal antibodies detected six months after vaccination in comparison to that of their corresponding plasmablast-derived monoclonal antibodies. Our data corroborate multiple reports demonstrating the maturation of circulating MBC responses after SARS-CoV-2 mRNA vaccination in humans<sup>[9,10,12,22,23,24](#)</sup>.

This study shows that a persistent vaccine-induced GC response in humans culminates in the induction of affinity-matured, antigen-specific BMPCs. Notably, none of the 11 bone marrow specimens came from a participant with a history of SARS-CoV-2 infection. An intriguing finding in our study is that the S-specific BMPCs detected more than six months after vaccination exhibited high SHM frequencies relative to other B cell compartments. These data corroborate similar observations made in the mouse model<sup>[25,26](#)</sup>. The mouse data led to the proposal that there is a division of labour between MBCs and long-lived BMPCs<sup>[27,28](#)</sup>. In that model, BMPCs secrete highly specific, high-affinity antibodies that provide the first layer of protection against the invading pathogen upon re-exposure, whereas MBCs would only be engaged in the event that the pathogen is not fully neutralized by BMPC-derived antibodies. Consistent with this notion, multiple reports have documented the evolution of circulating MBCs induced by SARS-CoV-2 mRNA vaccination in humans<sup>[9,10,12,23](#)</sup>. These reports have shown that not only did the frequency of circulating S-binding MBCs increased over time, but their ability to recognize S proteins from emerging SARS-CoV-2 variants seems to have expanded as well<sup>[22,23](#)</sup>. These data indicate an important role for affinity maturation of responding B cell clones beyond increasing binding affinity to the immunizing antigen.

Our study raises a number of important questions that will need to be addressed in future studies concerning the effects of an additional homologous or heterologous immunization on the dynamics and products of ongoing GCs, particularly with respect to breadth of induced B cell responses. It also remains to be addressed whether the IgA<sup>+</sup> GC B cell compartment induced by this systemic immunization can give rise to long-term IgA<sup>+</sup> MBCs and BMPCs. Overall, our data demonstrate the remarkable capacity of mRNA-based vaccines to induce robust and persistent GC reactions that culminate in affinity-matured MBC and BMPC populations.

## Methods

### Sample collection, preparation and storage

All studies were approved by the Institutional Review Board of Washington University in St Louis. Written consent was obtained from all participants. Forty-three healthy volunteers were enrolled, of whom 13 had a history of confirmed SARS-CoV-2

infection (Extended Data Table 1). Fifteen out of 43 healthy participants provided FNAs of draining axillary lymph nodes. In 6 out of the 15 participants, a second draining lymph node was identified and sampled following secondary immunization. One participant (15) received the boost vaccination in the contralateral arm; draining lymph nodes were identified and sampled on both sides. Eleven out of 43 healthy participants provided bone marrow aspirates. Forty-eight participants who had recovered from mild SARS-CoV-2 infection but had not been vaccinated within 7 months of illness were previously described<sup>21</sup>.

Peripheral blood samples were collected in EDTA tubes, and PBMCs were enriched by density gradient centrifugation over Ficoll-Paque PLUS (Cytiva) or Lymphopure (BioLegend). The residual red blood cells were lysed with ammonium chloride lysis buffer, and cells were immediately used or cryopreserved in 10% dimethyl sulfoxide in fetal bovine serum (FBS).

Ultrasound-guided FNA of draining axillary lymph nodes was performed by a radiologist or a qualified physician's assistant under the supervision of a radiologist. Scans were performed with a commercially available ultrasound unit (Logic E10, General Electric) using an L2–9 linear array transducer with transmit frequencies of 7, 8, and 9 MHz or a L6–15 linear array transducer with transmit frequencies of 10, 12, and 15 MHz. Lymph node dimensions and cortical thickness were measured, and the presence and degree of cortical vascularity and location of the lymph node relative to the axillary vein were determined before each FNA. For each FNA sample, six passes were made under continuous real-time ultrasound guidance using 25-gauge needles, each of which was flushed with 3 ml of RPMI 1640 supplemented with 10% FBS and  $100 \text{ U ml}^{-1}$  penicillin-streptomycin, followed by three 1-ml rinses. Red blood cells were lysed with ammonium chloride buffer (Lonza), washed with washing buffer (phosphate-buffered saline supplemented with 2% FBS and 2 mM EDTA), and immediately used or cryopreserved in 10% dimethyl sulfoxide in FBS. Participants reported no adverse effects from phlebotomies or serial FNAs.

Bone marrow aspirates of approximately 30 ml were collected in EDTA tubes from the iliac crest. Bone marrow mononuclear cells (BMMCs) were enriched by density gradient centrifugation over Ficoll-Paque PLUS, and then the remaining red blood cells were lysed with ammonium chloride buffer (Lonza) and washed with washing buffer. BMPCs were enriched from BMMCs using EasySep Human CD138 Positive Selection Kit II (StemCell Technologies) and immediately used for ELISpot or cryopreserved in 10% dimethyl sulfoxide in FBS.

## Antigens

Recombinant soluble S protein derived from SARS-CoV-2 was expressed as previously described<sup>29</sup>. In brief, a mammalian cell codon-optimized nucleotide sequences coding for the soluble version of S (GenBank: MN908947.3, amino acids 1–1213) including a C-terminal thrombin cleavage site, T4 fold trimerization domain and hexahistidine tag was cloned into the mammalian expression vector pCAGGS. The S protein sequence was modified to remove the polybasic cleavage site (RRAR to A) and two stabilizing mutations were introduced (K986P and V987P, wild-type numbering). Recombinant proteins were produced in Expi293F cells (Thermo Fisher Scientific) by transfection with purified plasmid using the ExpiFectamine 293 Transfection Kit (Thermo Fisher Scientific). Supernatants from transfected cells were collected 3 days after transfection, and recombinant proteins were purified using Ni-NTA agarose (Thermo Fisher Scientific), then buffer-exchanged into PBS and concentrated using Amicon Ultra centrifugal filters (MilliporeSigma). For flow cytometry staining, recombinant S was labeled with Alexa Fluor 7647-NHS ester or biotinylated using the EZ-Link Micro NHS-PEG4-Biotinylation Kit (Thermo Fisher Scientific); excess Alexa Fluor 647 and biotin were removed using 7-kDa Zeba desalting columns (Thermo Fisher Scientific).

For expression of biotinylated SARS-CoV-2 S Avitag, the CDS of pCAGGS vector containing recombinant soluble SARS-CoV-2 S protein was modified to encode 3' Avitag insert after the 6×His tag (5'-His tag-

GGCTCCGGGCTGAACGACATCTCGAAGCCCAGAAGATTGA  
GTGGCATGAG-Stop-3'; HHHHHHGSQLNDIFEAQKIEWHE-) using inverse PCR mutagenesis in a method described previously<sup>30</sup>. Protein expression and purification of SARS-CoV-2 S-Avitag was performed using the same methods as above. Immediately, after purification, site-specific biotinylation was performed similar to Avidity recommendations. Specifically, SARS-CoV-2 S-Avitag substrate was at 40 μM concentration with 15 μg ml<sup>-1</sup> BirA enzyme in a 0.05 M bicine buffer at pH 8.3 containing 10 mM ATP, 10 mM MgOAc and 50 μM biotin, and the reaction was performed for 30 °C for 1 h. The protein was then concentrated and buffer exchanged with PBS using a 100-kDa Amicon Ultra centrifugal filter (MilliporeSigma).

## Flow cytometry and cell sorting

Staining for flow cytometry analysis and sorting was performed using freshly isolated or cryo-preserved PBMCs or FNAs. For FNA staining, cells were incubated for 30 min on ice with biotinylated and Alexa Fluor 647-conjugated recombinant soluble S and PD-1-BB515 (EH12.1, BD Horizon, 1:100) in 2% FBS and 2 mM EDTA in PBS (P2), washed twice, then stained for 30 min on ice with IgG-BV480 (goat polyclonal, Jackson ImmunoResearch, 1:100), IgA-FITC (M24A, Millipore, 1:500), CD45-A532 (HI30, Thermo, 1:50), CD38-BB700 (HIT2, BD Horizon, 1:500), CD20-Pacific Blue (2H7, 1:400), CD27-BV510 (O323, 1:50), CD8-BV570 (RPA-T8, 1:200), IgM-BV605

(MHM-88, 1:100), HLA-DR-BV650 (L243, 1:100), CD19-BV750 (HIB19, 1:100), CXCR5-PE-Dazzle 594 (J252D4, 1:50), IgD-PE-Cy5 (IA6-2, 1:200), CD14-PerCP (HCD14, 1:50), CD71-PE-Cy7 (CY1G4, 1:400), CD4-Spark685 (SK3, 1:200), streptavidin-APC-Fire750, CD3-APC-Fire810 (SK7, 1:50) and Zombie NIR (all BioLegend) diluted in Brilliant Staining buffer (BD Horizon). Cells were washed twice with P2, fixed for 1 h at 25 °C using the True Nuclear fixation kit (BioLegend), washed twice with True Nuclear Permeabilization/Wash buffer, stained with FOXP3-BV421 (206D, BioLegend, 1:15), Ki-67-BV711 (Ki-67, BioLegend, 1:200), T-bet-BV785 (4B10, BioLegend, 1:400), BCL6-PE (K112-91, BD Pharmingen, 1:25), and BLIMP1-A700 (646702, R&D, 1:50) for 1 h at 25 °C, washed twice with True Nuclear Permeabilization/Wash buffer and resuspended in P2 for acquisition. For memory B cell staining, PBMC were incubated for 30 min on ice with biotinylated and Alexa Fluor 647-conjugated recombinant soluble S in P2, washed twice, then stained for 30 min on ice with IgG-BV480 (goat polyclonal, Jackson ImmunoResearch, 1:100), IgD-Super Bright 702 (IA6-2, Thermo, 1:50), IgA-FITC (M24A, Millipore, 1:500), CD45-A532 (HI30, Thermo, 1:50), CD38-BB700 (HIT2, BD Horizon, 1:500), CD24-BV421 (ML5, 1:100), CD20-Pacific Blue (2H7, 1:400), CD27-BV510 (O323, 1:50), CD8-BV570 (RPA-T8, 1:200), IgM-BV605 (MHM-88, 1:100), CD19-BV750 (HIB19, 1:100), FcRL5-PE (509f6, 1:100), CXCR5-PE-Dazzle 594 (J252D4, 1:50), CD14-PerCP (HCD14, 1:50), CD71-PE-Cy7 (CY1G4, 1:400), CD4-Spark685 (SK3, 1:200), streptavidin-APC-Fire750, CD3-APC-Fire810 (SK7, 1:50) and Zombie NIR (all BioLegend) diluted in Brilliant Staining buffer (BD Horizon). Cells were washed twice with P2 and resuspended in P2 for acquisition. All samples were acquired on an Aurora using SpectroFlo v.2.2 (Cytek). Flow cytometry data were analysed using FlowJo v.10 (BD Biosciences).

For sorting plasmablasts from peripheral blood, B cells were enriched from PBMC by first using EasySep Human Pan-B cell Enrichment Kit (StemCell Technologies), and then stained with CD20-PB (2H7, 1:400), CD3-FITC (HIT3a, 1:200), IgD-PerCP-Cy5.5 (IA6-2, 1:200), CD71-PE (CY1G4, 1:400), CD38-PE-Cy7 (HIT2, 1:200), CD19-APC (HIB19, 1:200) and Zombie Aqua (all BioLegend). For sorting GC B cells and LNPsCs from the lymph node, single-cell suspensions were stained for 30min on ice with PD-1-BB515 (EH12.1, BD Horizon, 1:100), CD20-Pacific Blue (2H7, 1:100), IgD-PerCP-Cy5.5 (IA6-2, 1:200), CD19-PE (HIB19, 1:200), CXCR5-PE-Dazzle 594 (J252D4, 1:50), CD38-PE-Cy7 (HIT2, 1:200), CD4-Alexa-Fluor-700 (SK3, 1:400), CD71-APC (CY1G4, 1:100), and Zombie Aqua (all BioLegend). Cells were washed twice, and single plasmablasts (live singlet CD19<sup>+</sup>CD3<sup>-</sup>IgD<sup>low</sup>CD38<sup>+</sup>CD20<sup>-</sup>CD71<sup>+</sup>), GC B cells (live singlet CD19<sup>+</sup>CD4<sup>-</sup>IgD<sup>low</sup>CD71<sup>+</sup>CD38<sup>int</sup>CD20<sup>+</sup>CXCR5<sup>+</sup>), LNPsCs (live singlet CD19<sup>+</sup>CD4<sup>-</sup>IgD<sup>low</sup>CD38<sup>+</sup>CD20<sup>-</sup>CD71<sup>+</sup>) were sorted using a FACSAria II.

## ELISA

Assays were performed in MaxiSorp 96-well plates (Thermo Fisher) coated with 100 µl of recombinant SARS-CoV-2 S, Donkey anti-human IgG (H+L) antibody (Jackson ImmunoResearch, 709-005-149) or BSA diluted to 1 µg ml<sup>-1</sup> in PBS, and plates were incubated at 4 °C overnight. Plates then were blocked with 10% FBS and 0.05% Tween 20 in PBS. Plasma or purified monoclonal antibodies were serially diluted in blocking buffer and added to the plates. Monoclonal antibodies and plasma samples were tested at 10 µg ml<sup>-1</sup> and 1:30 starting dilution, respectively, followed by 7 additional threefold serial dilutions. Plates were incubated for 90 min at room temperature and then washed 3 times with 0.05% Tween 20 in PBS. Secondary antibodies were diluted in blocking buffer before adding to wells and incubating for 60 min at room temperature. Horseradish peroxidase (HRP)-conjugated goat anti-human IgG (H+L) antibody (Jackson ImmunoResearch, 109-035-088, 1:2,500) was used to detect monoclonal antibodies. HRP-conjugated goat anti-Human IgG Fcγ fragment (Jackson ImmunoResearch, 109-035-190, 1:1,500), HRP-conjugated goat anti-human serum IgA α chain (Jackson ImmunoResearch, 109-035-011, 1:2,500), and HRP-conjugated goat anti-human IgM (Caltag, H15007, 1:4,000) were used to detect plasma antibodies. Plates were washed three times with PBST and three times with PBS before the addition of *O*-phenylenediamine dihydrochloride peroxidase substrate (MilliporeSigma). Reactions were stopped by the addition of 1 M hydrochloric acid. Optical density measurements were taken at 490 nm. The threshold of positivity for recombinant monoclonal antibodies was set as two times the optical density of background binding to BSA at the highest concentration of each monoclonal antibody. The area under the curve for each monoclonal antibody and half-maximal binding dilution for each plasma sample were calculated using GraphPad Prism v.9. Plasma antibody avidity was measured as previously described<sup>31</sup>. Areas under the curve were calculated by setting the mean + 3 × s.d. of background binding to BSA as a baseline. In brief, plasma dilutions that would give an optical density reading of 2.5 were calculated from the serial dilution ELISA. S-coated plates were incubated with this plasma dilution as above and then washed one time for 5 min with either PBS or 8 M urea in PBS, followed by 3 washes with PBST and developed as above. The avidity index was calculated for each sample as the optical density ratio of the urea-washed to PBS-washed wells.

## ELISpot

ELISpot plates were coated overnight at 4 °C with Flucelvax Quadrivalent 2019/2020 seasonal influenza virus vaccine (Seqirus, 1:100), tetanus/diphtheria vaccine (Grifols, 1:20), SARS-CoV-2 S (10 µg ml<sup>-1</sup>), anti-human Ig (Cellular Technology) and BSA. A direct ex vivo ELISpot assay was performed to determine the number of total, vaccine-binding or recombinant S-binding IgG- and IgA-secreting cells present in PBMCs or enriched BMPCs using Human IgA/IgG double-colour ELISpot kits (Cellular

Technology) according to the manufacturer's protocol. ELISpot plates were analysed using an ELISpot analyser (Cellular Technology).

## Single-cell RNA-seq library preparation and sequencing

Sorted plasmablasts and whole FNA from each time point were processed using the following 10x Genomics kits: Chromium Next GEM Single Cell 5' Kit v2 (PN-1000263); Chromium Next GEM Chip K Single Cell Kit (PN-1000286); BCR Amplification Kit (PN-1000253); Dual Index Kit TT Set A (PN-1000215). Chromium Single Cell 5' Gene Expression Dual Index libraries and Chromium Single Cell V(D)J Dual Index libraries were prepared according to manufacturer's instructions without modifications. Both gene expression and V(D)J libraries were sequenced on a Novaseq S4 (Illumina), targeting a median sequencing depth of 50,000 and 5,000 read pairs per cell, respectively.

## Bulk BCR sequencing

Sorted GC B cells and LNPCs from FNA, enriched BMPCs from bone marrow or enriched MBCs from PBMCs from blood were used for library preparation for bulk BCR sequencing. Circulating MBCs were magnetically isolated by first staining with IgD-PE and MojoSort anti-PE Nanobeads (BioLegend), and then processing with the EasySep Human B Cell Isolation Kit (StemCell Technologies) to negatively enrich IgDlo B cells. RNA was prepared from each sample using the RNeasy Plus Micro kit (Qiagen). Libraries were prepared using the NEBNext Immune Sequencing Kit for Human (New England Biolabs) according to the manufacturer's instructions without modifications. High-throughput  $2 \times 300$ -bp paired-end sequencing was performed on the Illumina MiSeq platform with a 30% PhiX spike-in according to manufacturer's recommendations, except for performing 325 cycles for read 1 and 275 cycles for read 2.

## Preprocessing of bulk sequencing BCR reads

Preprocessing of demultiplexed pair-end reads were performed using pRESTO v.0.6.2<sup>32</sup> as previously described<sup>6</sup>, with the exception that sequencing errors were corrected using the unique molecular identifiers (UMIs) as they were without additional clustering (Extended Data Table 5). Previously preprocessed unique consensus sequences from reported samples<sup>6</sup> were included as they were. Previously preprocessed unique consensus sequences from reported samples<sup>17</sup> corresponding to participants 01, 02a, 04, 07, 10, 13, 20, and 22 were subset to those with at least two contributing reads and included.

## Preprocessing of 10x Genomics single-cell BCR reads

Demultiplexed pair-end FASTQ reads were preprocessed using the ‘cellranger vdj’ command from 10x Genomics’ Cell Ranger v.6.0.1 for alignment against the GRCh38 human reference v.5.0.0 (‘refdata-cellranger-vdj-GRCh38-alts-ensembl-5.0.0’). The resultant ‘filtered\_contig.fasta’ files were used as preprocessed single-cell BCR reads (Extended Data Table 5).

## V(D)J gene annotation and genotyping

Initial germline V(D)J gene annotation was performed on the preprocessed BCRs using IgBLAST v.1.17.1<sup>33</sup> with the deduplicated version of IMGT/V-QUEST reference directory release 202113-2<sup>34</sup>. IgBLAST output was parsed using MakeDb.py from Change-O v.1.0.2<sup>35</sup>. For the single-cell BCRs, isotype annotation was pulled from the ‘c\_call’ column in the ‘filtered\_contig\_annotations.csv’ files outputted by Cell Ranger.

For both bulk and single-cell BCRs, sequence-level quality control was performed, requiring each sequence to have non-empty V and J gene annotations; exhibit chain consistency in all annotations; bear fewer than 10 non-informative (non-A/T/G/C, such as N or -) positions; and carry a non-empty CDR3 with no N and a nucleotide length that is a multiple of 3. For single-cell BCRs, cell-level quality control was also performed, requiring each cell to have either exactly one heavy chain and at least one light chain, or at least one heavy chain and exactly one light chain. Within a cell, for the chain type with more than one sequence, the most abundant sequence in terms of UMI count (when tied, the sequence that appeared earlier in the file) was kept. Ultimately, exactly one heavy chain and one light chain per cell were kept. Additionally, quality control against cross-sample contamination was performed by examining the extent, if any, of pairwise overlapping between samples in terms of BCRs with both identical UMIs and identical non-UMI nucleotide sequences.

Individualized genotypes were inferred based on sequences that passed all quality control using TIgGER v.1.0.0<sup>36</sup> and used to finalize V(D)J annotations. Sequences annotated as non-productively rearranged by IgBLAST were removed from further analysis.

## Clonal lineage inference

B cell clonal lineages were inferred on a by-individual basis based on productively rearranged sequences using hierarchical clustering with single linkage<sup>37</sup>. When combining both bulk and single-cell BCRs, heavy chain-based clonal inference was performed<sup>38</sup>. First, heavy chain sequences were partitioned based on common V and J

gene annotations and CDR3 lengths using the groupGenes function from Alakazam v1.1.0<sup>35</sup>. Within each partition, heavy chain sequences with CDR3s that were within 0.15 normalized Hamming distance from each other were clustered as clones using the hclust function from fastcluster v1.2.3<sup>39</sup>. When using only single-cell BCRs, clonal inference was performed based on paired heavy and light chains. First, paired heavy and light chains were partitioned based on common V and J gene annotations and CDR3 lengths. Within each partition, pairs whose heavy chain CDR3s were within 0.15 normalized Hamming distance from each other were clustered as clones.

Following clonal inference, full-length clonal consensus germline sequences were reconstructed using CreateGermlines.py from Change-O v.1.0.2<sup>35</sup> for each clone with the D-segment (for heavy chains) and the N/P regions masked with Ns, resolving any ambiguous gene assignments by majority rule. Within each clone, duplicate IMGT-aligned V(D)J sequences from bulk sequencing were collapsed using the collapseDuplicates function from Alakazam v1.1.0<sup>35</sup> except for duplicates derived from different time points, tissues, B cell compartments, or isotypes.

## BCR analysis

BCR analysis was performed in R v.4.1.0 with visualization performed using base R, ggplot2 v.3.3.5<sup>40</sup>, and GraphPad Prism v.9.

For the B cell compartment label, gene expression-based cluster annotation was used for single-cell BCRs; FACS-based sorting was used in general for bulk BCRs, except that plasmablast sorts from lymph nodes were labelled LNPCs, week 5 IgDlo sorts from blood were labelled activated, and week 7 IgDlo sorts from blood were labelled memory. For the time point label, one blood plasmablast sample that pooled collections in both week 4 and week 5 was treated as week 4; and one blood memory sort sample that pooled collections in both week 29 and week 30 was treated as week 29. For analysis involving the memory compartment, the memory sequences were restricted to bulk-sequenced week 29 memory sorts from blood.

A heavy chain-based B cell clone was considered a S-specific clone if the clone contained any sequence corresponding to a recombinant monoclonal antibody that was synthesized based on the single-cell BCRs and that tested positive for S-binding.

Clonal overlap between B cell compartments was visualized using circlize v.0.4.13<sup>41</sup>.

SHM frequency was calculated for each heavy chain sequence by counting the number of nucleotide mismatches from the germline sequence in the variable segment leading up to the CDR3, while excluding the first 18 positions that could be error-prone due to

the primers used for generating the monoclonal antibody sequences. Calculation was performed using the calcObservedMutations function from SHazaM v.1.0.2<sup>35</sup>.

Phylogenetic trees for S-specific clones containing BMPCs were constructed on a by-participant basis using IgPhyML v1.1.3<sup>18</sup> with the HLP19 model<sup>42</sup>. Only heavy chain sequences from week 4 plasmablast compartment, the GC B cell, LNPC and MBC compartments up to and including week 15, and the week 29 or 40 BMPC compartment were considered. For clones with > 100 sequences, subsampling was applied with probabilities proportional to the proportions of sequences from different compartments, in addition to keeping all sequences corresponding to synthesized monoclonal antibodies and all BMPC sequences. Only subsampled sequences from the plasmablast, LNPC and BMPC compartments were used for eventual tree-building. Trees were visualized using ggtree v3.0.4<sup>43</sup>.

## Human housekeeping genes

A list of human housekeeping genes was compiled from the 20 most stably expressed genes across 52 tissues and cell types in the Housekeeping and Reference Transcript (HRT) Atlas v.1.0<sup>44</sup>; 11 highly uniform and strongly expressed genes reported<sup>45</sup>; and some of the most commonly used housekeeping genes<sup>46</sup>. The final list includes 34 genes: *ACTB*, *TLE5* (also known as *AES*), *AP2M1*, *BSG*, *C1orf43*, *CD59*, *CHMP2A*, *CSNK2B*, *EDF1*, *EEF2*, *EMC7*, *GABARAP*, *GAPDH*, *GPI*, *GUSB*, *HNRNPA2B1*, *HPRT1*, *HSP90AB1*, *MLF2*, *MRFAP1*, *PCBP1*, *PFDN5*, *PSAP*, *PSMB2*, *PSMB4*, *RAB11B*, *RAB1B*, *RAB7A*, *REEP5*, *RHOA*, *SNRPD3*, *UBC*, *VCP* and *VPS29*.

## Processing of 10x Genomics single-cell 5' gene expression data

Demultiplexed pair-end FASTQ reads were first preprocessed on a by-sample basis using the ‘cellranger count’ command from 10× Genomics’ Cell Ranger v.6.0.1 for alignment against the GRCh38 human reference v.2020-A (‘refdata-gex-GRCh38-2020-A’). To avoid a batch effect introduced by sequencing depth, the ‘cellranger aggr’ command was used to subsample from each sample so that all samples had the same effective sequencing depth, which was measured in terms of the number of reads confidently mapped to the transcriptome or assigned to the feature IDs per cell. Gene annotation on human reference chromosomes and scaffolds in Gene Transfer Format (‘gencode.v32.primary\_assembly.annotation.gtf’) was downloaded (on 2 June 2021) from GENCODE v.32<sup>47</sup>, from which a biotype (‘gene\_type’) was extracted for each feature. Quality control was performed as follows on the aggregate gene expression matrix consisting of 432,713 cells and 36,601 features using SCANPY v.1.7.2<sup>48</sup> and Python v.3.8.8. (1) To remove presumably lysed cells, cells with mitochondrial content greater than 12.5% of all transcripts were removed. (2) To remove likely doublets, cells with more than 8,000 features or 80,000 total UMIs were removed. (3) To

remove cells with no detectable expression of common endogenous genes, cells with no transcript for any of the 34 housekeeping genes were removed. (4) The feature matrix was subset, based on their biotypes, to protein-coding, immunoglobulin, and T cell receptor genes that were expressed in at least 0.1% of the cells in any sample. The resultant feature matrix contained 15,842 genes. (5) Cells with detectable expression of fewer than 200 genes were removed. After quality control, there were a total of 383,708 cells from 56 single-cell samples (Extended Data Table 5).

## Single-cell gene expression analysis

Single-cell gene expression analysis was performed in SCANPY v.1.7.2<sup>48</sup>. UMI counts measuring gene expression were log-normalized. The top 2,500 highly variable genes (HVGs) were identified using the ‘scanpy.pp.highly\_variable\_genes’ function with the ‘seurat\_v3’ method, from which immunoglobulin and T cell receptor genes were removed. The data were scaled and centred, and principal component analysis (PCA) was performed based on HVG expression. PCA-guided neighborhood graphs embedded in uniform manifold approximation and projection (UMAP) were generated using the top 20 principal components via the ‘scanpy.pp.neighbors’ and ‘scanpy.tl.umap’ functions.

Overall clusters (Extended Data Table 3, top) were identified using Leiden graph-clustering via the ‘scanpy.tl.leiden’ function with resolution 0.23 (Extended Data Fig. 2b). UMAPs were faceted by batch, by participant, and by participant followed by sample; and inspected for convergence across batches, participants, and samples within participants, to assess whether there was a need for integration (Extended Data Fig. 2b). Cluster identities were assigned by examining the expression of a set of marker genes for different cell types (Extended Data Fig. 2c): *MS4A1*, *CD19* and *CD79A* for B cells; *CD3D*, *CD3E*, *CD3G*, *IL7R* and *CD4* or *CD8A* for CD4<sup>+</sup> or CD8<sup>+</sup> T cells, respectively; *GZMB*, *GNLY*, *NKG7* and *NCAM1* for NK cells; *CD14*, *LYZ*, *CST3* and *MS4A7* for monocytes; *IL3RA* and *CLEC4C* for plasmacytoid dendritic cells (pDCs); and *FDCSP*, *CXCL14*<sup>15</sup> and *FCAMR*<sup>16</sup> for FDCs. One group of 27 cells labelled ‘B and T’ was excluded. To remove potential contamination by platelets, 73 cells with a log-normalized expression value of >2.5 for PPBP were removed. All 644 cells from the FDC cluster were confirmed to have originated from FNA samples instead of blood.

Cells from the overall B cell cluster (Extended Data Table 3, bottom) were further clustered to identify B cell subsets using Leiden graph-clustering via the ‘scanpy.tl.leiden’ function with resolution 0.18 (Extended Data Fig. 2d). Cluster identities were assigned by examining the expression of a set of marker genes for different B cell subsets (Extended Data Fig. 2e) along with the availability of BCRs. The following marker genes were examined: *BCL6*, *RGS13*, *MEF2B*, *STMN1*, *ELL3*

and *SERPINA9* for GC B cells; *XBP1*, *IRF4*, *SEC11C*, *FKBP11*, *JCHAIN* and *PRDM1* for plasmablasts and LNPCs; *TCL1A*, *IL4R*, *CCR7*, *IGHM*, and *IGHD* for naive B cells; and *TNFRSF13B*, *CD27* and *CD24* for MBCs. Although one group clustered with B cells during overall clustering, it was labelled ‘B and T’ as its cells tended to have both BCRs and relatively high expression levels of *CD2* and *CD3E*; and was subsequently excluded from the final B cell clustering. Eighteen cells that were found in the GC B cell cluster but came from blood samples were labelled ‘PB-like’<sup>13</sup>. Two hundred and twenty-three cells that were found in the plasmablast cluster but came from FNA samples were re-assigned as LNPCs. Forty cells that were found in the LNPC cluster but came from blood samples were re-assigned as plasmablasts. Heavy chain SHM frequency and isotype usage of the B cell subsets were assessed for consistency with expected values to further confirm their assigned identities.

### **Selection of single-cell BCRs from GC B cell or LNPC clusters for expression**

Single-cell gene expression analysis was performed using lymph node samples up to and including week 15 on a by-participant basis. Clonal inference was performed based on paired heavy and light chains from the same samples. From every clone with a clone size of more than 3 cells that contained cells from the GC B cell and/or LNPC clusters, one GC B cell or LNPC was selected. For selection, where a clone spanned both the GC B cell and LNPC compartments, and/or multiple time points, a compartment and a timepoint were first randomly selected. Within that clone, the cell with the highest heavy chain UMI count was then selected, breaking ties based on *IGHV* SHM frequency. In all selected cells, native pairing was preserved.

### **Selection of BCRs from S-specific BMPC clones for expression**

From each heavy chain-based S-specific clone containing both plasmablasts and BMPCs, where possible, one plasmablast heavy chain was selected, and, together with all BMPC heavy chains, were paired with the same light chain for expression. For the plasmablast heavy chain, if single-cell paired plasmablasts were available, the single-cell paired plasmablast whose *IGHV* mutation frequency was closest to the median mutation frequency of other single-cell paired plasmablasts in the same clone (breaking ties by UMI count), and whose light chain V gene, J gene, and CDR3 length (VJL) combination was consistent with the clonal majority, was used as the source. The natively paired light chain of the plasmablast from which the heavy chain was selected was used. In clones in which two plasmablasts had inconsistent light chain VJL combinations, both plasmablasts were used. Clones in which there was light chain uncertainty due to more than two plasmablasts or due to LNPCs were generally excluded.

## Curation of selected BCRs for expression

The selected BCRs were curated prior to synthesis. First, artificial gaps introduced under the IMGT unique numbering system<sup>49</sup> were removed from the IMGT-aligned observed V(D)J sequences. IMGT gaps were identified as positions containing in-frame triplet dots ('...') in the IMGT-aligned germline sequences. Second, any non-informative (non-A/T/G/C, such as N or -) positions in the observed sequences, with the exception of potential in-frame indels, were patched by the nucleotides at their corresponding germline positions. Third, if applicable, the 3' end of the observed sequences were trimmed so that the total nucleotide length would be a multiple of 3. Finally, potential in-frame indels were manually reviewed. For a given potential in-frame indel from a selected cell, its presence or lack thereof in the unselected cells from the same clone was considered. Barring strong indications that an in-frame indel was due to sequencing error rather than the incapability of the IMGT unique numbering system of capturing it, the in-frame indels were generally included in the final curated sequences.

## Transfection for recombinant monoclonal antibodies and Fab production

Selected pairs of heavy and light chain sequences were synthesized by GenScript and sequentially cloned into IgG1, Igκ/λ and Fab expression vectors. Heavy and light chain plasmids were co-transfected into Expi293F cells (Thermo Fisher Scientific) for recombinant monoclonal antibody production, followed by purification with protein A agarose resin (GoldBio). Expi293F cells were cultured in Expi293 Expression Medium (Gibco) according to the manufacturer's protocol.

## GFP-reduction neutralization test

Serial dilutions of each monoclonal antibody diluted in DMEM were incubated with  $10^2$  plaque-forming unit (PFU) of VSV-SARS-CoV-2 D614G for 1 h at 37 °C. Antibody-virus complexes were added to Vero cell monolayers in 96-well plates and incubated at 37 °C for 7.5 h. Cells were fixed at room temperature in 2% formaldehyde (Millipore Sigma) containing 10 µg ml<sup>-1</sup> of Hoechst 33342 nuclear stain (Invitrogen) for 45 min at room temperature. Fixative was replaced with PBS prior to imaging. Images were acquired using an IN Cell 2000 Analyzer automated microscope (GE Healthcare) in both the DAPI and FITC channels to visualize nuclei and infected cells. Images were analysed using the Multi Target Analysis Module of the IN Cell Analyzer 1000 Workstation Software (GE Healthcare). GFP-positive cells were identified using the top-hat segmentation method and subsequently counted within the IN Cell

workstation software. The initial dilution of monoclonal antibody started at 25 µg ml<sup>-1</sup> and was threefold serially diluted in 96-well plate over 8 dilutions.

## Affinity analysis by BLI

We used the Octet Red instrument (ForteBio) with shaking at 1,000 rpm. The kinetic analysis using Octet SA biosensors (Sartorius) was performed as follows: (1) baseline: 120 s immersion in buffer (10 mM HEPES and 1% BSA); (2) loading: 130 s immersion in solution with 10 µg ml<sup>-1</sup> biotinylated SARS-CoV-2 S Avitag; (3) baseline: 120 s immersion in buffer; (4) association: 300 s immersion in solution with serially diluted recombinant Fab; (5) dissociation: 600 s immersion in buffer. The BLI signal was recorded and analysed using BIAevaluation Software (Biacore). The 1:1 binding model with a drifting baseline was employed for the equilibrium dissociation constant ( $K_D$ ).

## Reporting summary

Further information on research design is available in the [Nature Research Reporting Summary](#) linked to this paper.

## Data availability

Raw sequencing data and transcriptomics count matrix are deposited at the Sequence Read Archive ([PRJNA777934](#)) and Gene Expression Omnibus ([PRJNA777934](#)). Processed transcriptomics and BCR data are deposited at Zenodo (<https://doi.org/10.5281/zenodo.5895181>). Previously reported bulk-sequenced BCR data used in this study were deposited at the Sequence Read Archive under PRJNA731610 and PRJNA741267, and at <https://doi.org/10.5281/zenodo.5042252> and <https://doi.org/10.5281/zenodo.5040099>.

## References

1. Victora, G. D. & Nussenzweig, M. C. Germinal centers. *Annu. Rev. Immunol.* **30**, 429–457 (2012).
2. Cyster, J. G. & Allen, C. D. C. B cell responses: cell interaction dynamics and decisions. *Cell* **177**, 524–540 (2019).
3. Radbruch, A. et al. Competence and competition: the challenge of becoming a long-lived plasma cell. *Nat. Rev. Immunol.* **6**, 741–750 (2006).

4. Slifka, M. K., Antia, R., Whitmire, J. K. & Ahmed, R. Humoral immunity due to long-lived plasma cells. *Immunity* **8**, 363–372 (1998).
5. Tarlinton, D., Radbruch, A., Hiepe, F. & Dörner, T. Plasma cell differentiation and survival. *Curr. Opin. Immunol.* **20**, 162–169 (2008).
6. Turner, J. S. et al. SARS-CoV-2 mRNA vaccines induce persistent human germinal centre responses. *Nature* **596**, 109–113 (2021).
7. Mudd, P. A. et al. SARS-CoV-2 mRNA vaccination elicits a robust and persistent T follicular helper cell response in humans. *Cell* **185**, 603–613.e15 (2021).
8. Laidlaw, B. J. & Ellebedy, A. H. The germinal centre B cell response to SARS-CoV-2. *Nat. Rev. Immunol.* **22**, 7–18 (2021).
9. Goel, R. R. et al. mRNA vaccines induce durable immune memory to SARS-CoV-2 and variants of concern. *Science* **374**, abm0829 (2021).
10. Cho, A. et al. Anti-SARS-CoV-2 receptor binding domain antibody evolution after mRNA vaccination. *Nature* **600**, 517–522 (2021).
11. Chen, Y. et al. Differential antibody dynamics to SARS-CoV-2 infection and vaccination. Preprint at <https://doi.org/10.1101/2021.09.09.459504> (2021).
12. Pape, K. A. et al. High-affinity memory B cells induced by SARS-CoV-2 infection produce more plasmablasts and atypical memory B cells than those primed by mRNA vaccines. *Cell Rep.* **37**, 109823 (2021).
13. Turner, J. S. et al. Human germinal centres engage memory and naive B cells after influenza vaccination. *Nature* **586**, 127–132 (2020).
14. King, H. W. et al. Single-cell analysis of human B cell maturation predicts how antibody class switching shapes selection dynamics. *Sci. Immunol.* **6**, eabe6291 (2021).
15. Haebe, S. et al. Single-cell analysis can define distinct evolution of tumor sites in follicular lymphoma. *Blood* **137**, 2869–2880 (2021).
16. Mourcin, F. et al. Follicular lymphoma triggers phenotypic and functional remodeling of the human lymphoid stromal cell landscape. *Immunity* **54**, 1901 (2021).
17. Schmitz, A. J. et al. A vaccine-induced public antibody protects against SARS-CoV-2 and emerging variants. *Immunity* **54**, 2159–2166.e2156 (2021).

18. Hoehn, K. B., Lunter, G. & Pybus, O. G. A phylogenetic codon substitution model for antibody lineages. *Genetics* **206**, 417–427 (2017).
19. Liu, Z. et al. Identification of SARS-CoV-2 spike mutations that attenuate monoclonal and serum antibody neutralization. *Cell Host Microbe* **29**, 477–488.e474 (2021).
20. Elsner, R. A. & Shlomchik, M. J. Germinal center and extrafollicular B cell responses in vaccination, immunity, and autoimmunity. *Immunity* **53**, 1136–1150 (2020).
21. Turner, J. S. et al. SARS-CoV-2 infection induces long-lived bone marrow plasma cells in humans. *Nature* **595**, 421–425 (2021).
22. Tong, P. et al. Memory B cell repertoire for recognition of evolving SARS-CoV-2 spike. *Cell* **184**, 4969–4980.e4915 (2021).
23. Sokal, A. et al. mRNA vaccination of naive and COVID-19-recovered individuals elicits potent memory B cells that recognize SARS-CoV-2 variants. *Immunity* **54**, 2893–2907.e5 (2021).
24. Lucas, C. et al. Impact of circulating SARS-CoV-2 variants on mRNA vaccine-induced immunity. *Nature* **600**, 523–529 (2021).
25. Smith, K. G. C. The extent of affinity maturation differs between the memory and antibody-forming cell compartments in the primary immune response. *EMBO J.* **16**, 2996–3006 (1997).
26. Weisel, F. J., Zuccarino-Catania, G. V., Chikina, M. & Shlomchik, M. J. A temporal switch in the germinal center determines differential output of memory B and plasma cells. *Immunity* **44**, 116–130 (2016).
27. Purtha, W. E., Tedder, T. F., Johnson, S., Bhattacharya, D. & Diamond, M. S. Memory B cells, but not long-lived plasma cells, possess antigen specificities for viral escape mutants. *J. Exp. Med.* **208**, 2599–2606 (2011).
28. Pape, K. A., Taylor, J. J., Maul, R. W., Gearhart, P. J. & Jenkins, M. K. Different B cell populations mediate early and late memory during an endogenous immune response. *Science* **331**, 1203–1207 (2011).
29. Stadlbauer, D. et al. SARS-CoV-2 seroconversion in humans: a detailed protocol for a serological assay, antigen production, and test setup. *Curr. Protoc. Microbiol.* **57**, e100 (2020).

30. Fairhead, M. & Howarth, M. Site-specific biotinylation of purified proteins using BirA. *Methods Mol. Biol.* **1266**, 171–184 (2015).
31. Davis, C. W. et al. Longitudinal analysis of the human B cell response to ebola virus infection. *Cell* **177**, 1566–1582.e1517 (2019).
32. Vander Heiden, J. A. et al. pRESTO: a toolkit for processing high-throughput sequencing raw reads of lymphocyte receptor repertoires. *Bioinformatics* **30**, 1930–1932 (2014).
33. Ye, J., Ma, N., Madden, T. L. & Ostell, J. M. IgBLAST: an immunoglobulin variable domain sequence analysis tool. *Nucleic Acids Res.* **41**, W34–W40 (2013).
34. Brochet, X., Lefranc, M. P. & Giudicelli, V. IMGT/V-QUEST: the highly customized and integrated system for IG and TR standardized V-J and V-D-J sequence analysis. *Nucleic Acids Res.* **36**, W503–W508 (2008).
35. Gupta, N. T. et al. Change-O: a toolkit for analyzing large-scale B cell immunoglobulin repertoire sequencing data: Table 1. *Bioinformatics* **31**, 3356–3358 (2015).
36. Gadala-Maria, D., Yaari, G., Uduman, M. & Kleinstein, S. H. Automated analysis of high-throughput B-cell sequencing data reveals a high frequency of novel immunoglobulin V gene segment alleles. *Proc. Natl Acad. Sci. USA* **112**, E862–E870 (2015).
37. Gupta, N. T. et al. Hierarchical clustering can identify B cell clones with high confidence in Ig repertoire sequencing data. *J. Immunol.* **198**, 2489–2499 (2017).
38. Zhou, J. Q. & Kleinstein, S. H. Cutting edge: Ig H chains are sufficient to determine most B cell clonal relationships. *J. Immunol.* **203**, 1687–1692 (2019).
39. Müllner, D. fastcluster: fast hierarchical, agglomerative clustering routines for R and Python. *J. Stat. Softw.* **53**, 1–18 (2013).
40. Wickham, H. *ggplot2: Elegant Graphics for Data Analysis* (Springer, 2016).
41. Gu, Z., Gu, L., Eils, R., Schlesner, M. & Brors, B. circlize implements and enhances circular visualization in R. *Bioinformatics* **30**, 2811–2812 (2014).
42. Hoehn, K. B. et al. Repertoire-wide phylogenetic models of B cell molecular evolution reveal evolutionary signatures of aging and vaccination. *Proc. Natl Acad. Sci. USA* **116**, 22664–22672 (2019).

43. Yu, G., Smith, D. K., Zhu, H., Guan, Y. & Lam, T. T. Y. ggtree: an r package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods Ecol. Evol.* **8**, 28–36 (2017).
44. Hounkpe, B. W., Chenou, F., Franciele, & Erich, HRT Atlas v1.0 database: redefining human and mouse housekeeping genes and candidate reference transcripts by mining massive RNA-seq datasets. *Nucleic Acids Res.* **49**, D947–D955 (2021).
45. Eisenberg, E. & Levanon, E. Y. Human housekeeping genes, revisited. *Trends Genet.* **29**, 569–574 (2013).
46. Valente, V. et al. Selection of suitable housekeeping genes for expression analysis in glioblastoma using quantitative RT–PCR. *BMC Mol. Biol.* **10**, 17 (2009).
47. Frankish, A. et al. GENCODE reference annotation for the human and mouse genomes. *Nucleic Acids Res.* **47**, D766–D773 (2019).
48. Wolf, F. A., Angerer, P. & Theis, F. J. SCANPY: large-scale single-cell gene expression data analysis. *Genome Biol.* **19**, 15 (2018).
49. Lefranc, M.-P. et al. IMGT unique numbering for immunoglobulin and T cell receptor variable domains and Ig superfamily V-like domains. *Dev. Comp. Immunol.* **27**, 55–77 (2003).

## Acknowledgements

We thank the study participants for providing these specimens. We thank B. J. Laidlaw for critical reading of the manuscript and C. A. Nelson for help with the binding affinity analyses. The Ellebedy laboratory was supported by National Institute of Allergy and Infectious Diseases (NIAID) grants U01AI141990 and 1U01AI150747, NIAID Centers of Excellence for Influenza Research and Surveillance contract HHSN272201400006C and HHSN272201400008C, and NIAID Collaborative Influenza Vaccine Innovation Centers contract 75N93019C00051. W.K. was supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (2021R1A6A3A03041509). J.S.T. was supported by NIAID 5T32CA009547. This study used samples obtained from the Washington University School of Medicine’s COVID-19 biorepository supported by the NIH/National Center for Advancing Translational Sciences, grant number UL1 TR002345. The content of this manuscript is solely the responsibility of the authors and does not necessarily represent the official view of NIAID or NIH. The WU353 and WU368 studies were reviewed and approved by the Washington University

Institutional Review Board (approval numbers 202003186 and 202012081, respectively).

## Author information

### Author notes

1. These authors contributed equally: Wooseob Kim, Julian Q. Zhou

### Affiliations

1. Department of Pathology and Immunology, Washington University School of Medicine, St Louis, MO, USA

Wooseob Kim, Julian Q. Zhou, Stephen C. Horvath, Aaron J. Schmitz, Alexandria J. Sturtz, Tingting Lei, Mahima Thapa, Wafaa B. Alsoussi, Jackson S. Turner & Ali H. Ellebedy

2. Department of Molecular Microbiology, Washington University School of Medicine, St Louis, MO, USA

Zhuoming Liu & Sean P. J. Whelan

3. Division of Allergy and Immunology, Department of Internal Medicine, Washington University School of Medicine, St Louis, MO, USA

Elizaveta Kalaidina

4. Clinical Trials Unit, Washington University School of Medicine, St Louis, MO, USA

Alem Haile & Michael K. Klebert

5. Mallinckrodt Institute of Radiology, Washington University School of Medicine, St Louis, MO, USA

Teresa Suessen, William D. Middleton & Sharlene A. Teeffey

6. Division of Infectious Diseases, Department of Internal Medicine, Washington University School of Medicine, St Louis, MO, USA

Luis Parra-Rodriguez, Jane A. O'Halloran & Rachel M. Presti

7. Department of Emergency Medicine, Washington University School of Medicine, St Louis, MO, USA

Philip A. Mudd

8. Center for Vaccines and Immunity to Microbial Pathogens, Washington University School of Medicine, St Louis, MO, USA

Philip A. Mudd, Rachel M. Presti & Ali H. Ellebedy

9. Division of Oncology, Department of Medicine, Washington University School of Medicine, St Louis, MO, USA

Iskra Pusic

10. The Andrew M. and Jane M. Bursky Center for Human Immunology and Immunotherapy Programs, Washington University School of Medicine, St Louis, MO, USA

Ali H. Ellebedy

## Contributions

A.H.E. conceived and designed the study. A.H., M.K.K., L.P.-R., P.A.M., I.P., J.A.O. and R.M.P. wrote and maintained the IRB protocol, recruited, and phlebotomized participants, coordinated sample collection, and analysed clinical data. W.K., E.K. and J.S.T. processed specimens. W.K., E.K., W.B.A. and J.S.T. performed ELISA and ELISpot. W.K., S.C.H., A. J. Schmitz, T.L., M.T. and W.B.A. generated and characterized monoclonal antibodies. W.K. and A. J. Sturtz prepared libraries for scRNA sequencing. A. J. Schmitz performed RNA extractions and library preparation for bulk BCR sequencing. J.Q.Z. analysed scRNA sequencing and bulk BCR sequencing data. A. J. Schmitz expressed SARS-CoV-2 S proteins. J.S.T. sorted cells and collected and analysed the flow cytometry data. T.S. and W.D.M. performed FNA. W.D.M. and S.A.T. supervised lymph node evaluation prior to FNA and specimen collection and evaluated lymph node ultrasound data. Z.L. and S.P.J.W. performed and analysed the in vitro neutralization assay. W.K., J.Q.Z., J.S.T. and A.H.E. analysed the data. A.H.E. supervised experiments and obtained funding. W.K., J.Q.Z., P.A.M., J.S.T. and A.H.E. wrote the manuscript. All authors reviewed the manuscript.

## Corresponding author

Correspondence to [Ali H. Ellebedy](#).

## Ethics declarations

### Competing interests

The Ellebedy laboratory received funding under sponsored research agreements from Emergent BioSolutions and from AbbVie. J.S.T. is a consultant for Gerson Lehrman Group.

### Peer review

#### Peer review information

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

### Additional information

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

### Extended data figures and tables

#### [Extended Data Fig. 1 Persistence of humoral immune responses to SARS-CoV-2 mRNA vaccination.](#)

**a**, Flow cytometry gating strategies for GC B cells (Fig. [1b](#)) and LNPCs (defined as CD19<sup>+</sup> CD3<sup>-</sup> IgD<sup>low</sup> CD20<sup>low</sup> CD38<sup>+</sup> BLIMP1<sup>+</sup> CD71<sup>+</sup> live singlet lymphocytes) in the lymph node. **b**, Kinetics of total (left) and S-specific LNPCs (right) as gated in **a**. **c**, Frequencies of BMPCs secreting IgA antibodies specific for the indicated antigens 29 weeks after immunization. Symbols represent one sample in **b** ( $n = 15$ ) and **c** ( $n = 11$ ). **d**, **e**, Plasma antibody titers against SARS-CoV-2 S measured by ELISA in participants without (red,  $n = 29$ ) and with (black,  $n = 9$ ) a history of SARS-CoV-2 infection in SARS-CoV-2 vaccinated (left, center) and unvaccinated (right,  $n = 48$ ) participants 29 weeks after the first vaccine dose or symptom onset (**d**) and in vaccinated participants (red,  $n = 29$ ; black,  $n = 9$ ) over time (**e**).  $P$  values were determined by Kruskal-Wallis test followed by Dunn's multiple comparison test between unvaccinated and both vaccinated groups in **d**. Horizontal lines indicate median values and geometric mean, respectively, in **c** and **d**. Dotted lines indicate detection limit in **c**, **d** and **e**. **f**, Flow cytometry gating strategies for MBCs (CD19<sup>+</sup>

CD3<sup>-</sup> IgD<sup>low</sup> CD20<sup>+</sup> CD38<sup>-</sup> live singlet lymphocytes) and S-binding MBCs (Fig. 1g) in blood.

### Extended Data Fig. 2 Identification of SARS-CoV-2 S-binding B cell clones in the lymph node.

**a**, Flow cytometry gating strategies for sorting PBs (defined as CD19<sup>+</sup> CD3<sup>-</sup> IgD<sup>low</sup> CD20<sup>low</sup> CD38<sup>+</sup> CD71<sup>+</sup> live singlet lymphocytes) from blood. **b, d** UMAPs showing scRNA-seq transcriptional clusters of total cells (**b**) and of B cells (**d**) from PBs sorted from blood and FNA of draining axillary lymph nodes combined. **c, e**, Dot plots for the marker genes used for identifying annotated clusters. **f**, Heatmap of paired *IGHV* and *IGHJ* gene usage in S-binding clones. Color indicates the number of participants in which clones using a combination of *IGHV* and *IGHJ* genes were found. **g**, Flow cytometry gating strategies for sorting GC B cells (CD19<sup>+</sup> CD4<sup>-</sup> IgD<sup>low</sup> CD20<sup>+</sup> CD38<sup>int</sup> CXCR5<sup>high</sup> CD71<sup>+</sup> live singlet lymphocytes) and LNPCs (CD19<sup>+</sup> CD4<sup>-</sup> IgD<sup>low</sup> CD20<sup>low</sup> CD38<sup>+</sup> CXCR5<sup>low</sup> CD71<sup>+</sup> live singlet lymphocytes) from FNAs. **h**, SARS-CoV-2 S-binding clones visualized in red on UMAP of B cell clusters. Percentages are of S-binding clones within GC B cells (blue), LNPCs (green), PBs (red), MBCs (pink) or naive B cells (yellow). Total numbers of cells are at the bottom right corner.

### Extended Data Fig. 3 Maturation of SARS-CoV-2 S-binding B cells in the lymph node.

**a**, Circos diagrams showing clonal overlap between S-binding GC B cells at indicated time points. Purple and grey chords correspond to, respectively, clones spanning both 29 weeks post-vaccination and other time points, and clones spanning one or more of 4, 7 and 15 weeks post-vaccination. Percentages are of GC B cell clones related to GC B cells detected at 29 weeks post-vaccination. **b**, Circos diagrams showing clonal overlap between S-binding MBCs in blood 29 weeks post-vaccination and GC B cells at indicated time points. Purple and grey chords correspond to, respectively, clones spanning both the MBC and GC B cell compartments, and clones spanning only the GC B cell compartment. Percentages are of GC B cell clones overlapping with MBCs in blood 29 weeks post-vaccination. Arc length corresponds to the number of BCR sequences and chord width corresponds to clone size in **a** and **b**. **c**, Comparison of *IGHV* nucleotide mutation frequency of SARS-CoV-2 S-binding GC B cells in each participant at the indicated time points. Horizontal lines represent median values. Cell numbers are presented on the top of each data set. **d**, Comparison of *IGHV* region nucleotide mutation frequencies between clonally related, SARS-CoV-2 S-binding GC B cells and MBCs (n = 33) detected at 29 weeks post-vaccination. Each dot represents the median SHM frequency of a clone within the indicated compartment. Median

values are presented on the top of each data set.  $P$  value was determined by a paired two-sided non-parametric Mann-Whitney test. **e**, Percentages of GC B cells expressing BCRs of isotype IgG (blue), IgA (red), IgM (green) or IgD (pink) at the early (E) or the late (L) time point. The early and late time points represent, respectively, 4, 5 or 7 weeks, and 15 or 29 weeks after immunization. Cell numbers are at the top.

### **Extended Data Fig. 4 Evolution of B cell clones induced by SARS-CoV-2 vaccination.**

**a, a**, Comparison of *IGHV* nucleotide mutation frequency of PBs ( $n = 2735$ ), GC B cells ( $n = 139322$ ), LNPCs ( $n = 82350$ ), MBCs ( $n = 341$ ) and BMPCs ( $n = 47$ ). Horizontal lines represent median values.  $P$  values were determined by Kruskal-Wallis test followed by Dunn's multiple comparison test. **b**, Phylogenetic trees of neutralizing clones showing inferred evolutionary relationships between PBs (squares), LNPCs (triangles) and BMPCs (diamonds). Horizontal branch length represents the expected number of substitutions per codon in V-region genes, corresponding to the scale bar. Clone IDs are displayed near the root of the trees. Asterisks denote neutralizing mAbs. **c**, Kinetic curves of BLI signal for clonally related, PB- and BMPC-derived Fabs interacting with immobilized SARS-CoV-2 S. Clone IDs, Fab IDs and cell types are presented on the top of each data set. Asterisks denote neutralizing clones. **d**, Equilibrium dissociation constant ( $K_D$ ) of Fabs ( $n = 24$ ) interacting with immobilized SARS-CoV-2 S measured by biolayer interferometry (BLI). Red and black dots indicate  $K_D$  values of clonally related, PB- and BMPC-derived Fabs, respectively.  $P$  value was determined by Wilcoxon matched-pair signed rank test. **e**, Neutralization curves of VSV-SARS-CoV-2 D614G with BMPC-derived mAbs. Colored and grey lines represent neutralizing and non-neutralizing clones, respectively. Neutralizing clone IDs are indicated on each curve. ns > 0.9999, \*\*\* $P$  < 0.0001.

**Extended Data Table 1 Demographics of participants and vaccine side effects**

**Extended Data Table 2 Frequencies of GC B cells, LNPCs and CD14<sup>+</sup> myeloid cells in draining axillary lymph nodes**

**Extended Data Table 3 Cell counts and frequencies of transcriptional clusters and of SARS-CoV-2 S binding cells in scRNA-seq of PBs from blood and FNA from lymph nodes**

**Extended Data Table 4 Description of SARS-CoV-2 S-binding mAbs derived from GC B cells and LNPCs**

**Extended Data Table 5 Processing of BCR and 5' gene expression data from scRNA-seq and BCR reads from bulk-seq**

**Extended Data Table 6 Description of recombinant mAbs and Fabs derived from clonally related PBs and BMPCs**

# Supplementary information

## Reporting Summary

## Peer Review File

## Rights and permissions

### Reprints and Permissions

## About this article

### Cite this article

Kim, W., Zhou, J.Q., Horvath, S.C. *et al.* Germinal centre-driven maturation of B cell response to mRNA vaccination. *Nature* **604**, 141–145 (2022).  
<https://doi.org/10.1038/s41586-022-04527-1>

- Received: 31 October 2021
- Accepted: 04 February 2022
- Published: 15 February 2022
- Issue Date: 07 April 2022
- DOI: <https://doi.org/10.1038/s41586-022-04527-1>

## Share this article

Anyone you share the following link with will be able to read this content:

[Get shareable link](#)

Sorry, a shareable link is not currently available for this article.

[Copy to clipboard](#)

Provided by the Springer Nature SharedIt content-sharing initiative

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

[Download PDF](#)

- Article
- Open Access
- [Published: 30 March 2022](#)

# Genetic instability from a single S phase after whole-genome duplication

- [Simon Gemble](#) [ORCID: orcid.org/0000-0002-4351-7136<sup>1</sup>](#),
- [René Wardenaar](#) [ORCID: orcid.org/0000-0001-9891-1897<sup>2 na1</sup>](#),
- [Kristina Keuper](#) [ORCID: orcid.org/0000-0002-9841-5516<sup>3 na1</sup>](#),
- [Nishit Srivastava](#) [ORCID: orcid.org/0000-0003-4177-6123<sup>4</sup>](#),
- [Maddalena Nano<sup>1 nAff7</sup>](#),
- [Anne-Sophie Macé<sup>5</sup>](#),
- [Andréa E. Tijhuis<sup>2</sup>](#),
- [Sara Vanessa Bernhard<sup>3</sup>](#),
- [Diana C. J. Spierings](#) [ORCID: orcid.org/0000-0001-8403-474X<sup>2</sup>](#),
- [Anthony Simon<sup>1</sup>](#),
- [Oumou Goundiam<sup>1</sup>](#),
- [Helfrid Hochegger<sup>6</sup>](#),
- [Matthieu Piel](#) [ORCID: orcid.org/0000-0002-2848-177X<sup>4</sup>](#),
- [Floris Foijer](#) [ORCID: orcid.org/0000-0003-0989-3127<sup>2</sup>](#),
- [Zuzana Storchová](#) [ORCID: orcid.org/0000-0003-2376-7047<sup>3</sup>](#) &
- [Renata Basto](#) [ORCID: orcid.org/0000-0001-5556-2770<sup>1</sup>](#)

[Nature](#) volume 604, pages 146–151 (2022)

- 14k Accesses
- 1 Citations
- 125 Altmetric
- [Metrics details](#)

## Subjects

- [Cell growth](#)
- [Genomic instability](#)
- [Polyploidy](#)

## Abstract

Diploid and stable karyotypes are associated with health and fitness in animals. By contrast, whole-genome duplications—doublings of the entire complement of chromosomes—are linked to genetic instability and frequently found in human cancers<sup>1,2,3</sup>. It has been established that whole-genome duplications fuel chromosome instability through abnormal mitosis<sup>4,5,6,7,8</sup>; however, the immediate consequences of tetraploidy in the first interphase are not known. This is a key question because single whole-genome duplication events such as cytokinesis failure can promote tumorigenesis<sup>9</sup>. Here we find that human cells undergo high rates of DNA damage during DNA replication in the first S phase following induction of tetraploidy. Using DNA combing and single-cell sequencing, we show that DNA replication dynamics is perturbed, generating under- and over-replicated regions. Mechanistically, we find that these defects result from a shortage of proteins during the G1/S transition, which impairs the fidelity of DNA replication. This work shows that within a single interphase, unscheduled tetraploid cells can acquire highly abnormal karyotypes. These findings provide an explanation for the genetic instability landscape that favours tumorigenesis after tetraploidization.

## Main

As whole-genome duplications (WGDs) can have different origins<sup>10,11</sup>, we developed several approaches to induce tetraploidization through either mitotic slippage, cytokinesis failure or endoreplication in the diploid and genetically stable RPE-1 human cell line. Most cells resulting from cytokinesis failure contained two nuclei, whereas endoreplication or mitotic slippage generated mononucleated tetraploid cells. Cell size, cell number, nucleus size and centrosome number were considered to distinguish diploid cells from tetraploid cells (Fig. 1a, b, Extended Data Fig. 1a–i). For each approach, a mix of diploid and tetraploid cells was obtained, enabling the comparison of internal diploid controls and tetraploids. In all conditions, most tetraploid cells continued to cycle throughout the first interphase, allowing us to probe the consequences of tetraploidy within the first cell cycle.

**Fig. 1: High levels of DNA damage in the first interphase following unscheduled WGD.**

---

 **figure 1**

**a**, Schematic of the methods used to generate tetraploid cells. **b**, Top, images of diploid (D) and tetraploid (T) RPE-1 cells generated by mitotic slippage, cytokinesis failure or endoreplication. Centrosomes labelled with anti-CEP192 and cell membranes labelled with anti- $\beta$ -catenin. Bottom, outlined regions are shown at higher magnification. **c, e, g**, Images showing DNA damage caused by mitotic slippage (**c**), cytokinesis failure (**e**) or endoreplication (**g**) revealed by anti- $\gamma$ H2AX in diploid and tetraploid RPE-1 cells as indicated. **d, f, h**, The number of  $\gamma$ H2AX foci following mitotic slippage (**d**), cytokinesis failure (**f**) or endoreplication (**h**) per interphase cell in diploid and tetraploid RPE-1 cells. Data are mean  $\pm$  s.e.m.; >100 interphase cells, 3 independent experiments. The percentage of interphase cells with at least ten  $\gamma$ H2AX foci for each condition is indicated under the graph. **i, k**, Images of diploid and tetraploid RPE-1 cells generated by mitotic slippage labelled with anti-FANCD2 (**i**) or anti-53BP1 (**k**) antibodies. **j, l**, The number of FANCD2 (**j**) or 53BP1 (**l**) foci per interphase cell in diploid and tetraploid RPE-1 cells. Data are mean  $\pm$  s.e.m.; >100 interphase cells, 3 independent experiments. Dotted lines indicate the nuclear region. CF, cytokinesis failure; ENR, endoreplication; MS, mitotic slippage. **d, f, h, j, l**, One-sided analysis of variance (ANOVA) test. Scale bars, 10  $\mu$ m.

#### Source data

Using  $\gamma$ H2AX, an early marker of DNA damage, we found high levels of DNA damage in tetraploid cells (but not in controls) independently of how they were generated (Fig. [1c–h](#), Extended Data Figs. [1a–i](#), [2a–f](#), Methods). Moreover, whereas more than 10  $\gamma$ H2AX foci were present in only 5–9% of diploid cells, this proportion reaches 34–54% in tetraploid cells (Fig. [1c–h](#)). The number of  $\gamma$ H2AX foci correlated

with fluorescence intensity (Extended Data Fig. 1j). We excluded the possibility that the increase in tetraploid cells was simply owing to increased nuclear size by normalizing the number of  $\gamma$ H2AX foci to the nuclear area or nuclear fluorescence intensity (Extended Data Fig. 1k–l). High levels of DNA damage were also found in tetraploid BJ fibroblast and HCT116 cells upon WGD (Extended Data Fig. 2g, h).

To evaluate levels of DNA damage after WGD, we compared DNA damage between tetraploid and diploid cells with replication stress. Replication stress results from the slowing or stalling of replication forks, which can be induced by high doses of aphidicolin (APH; a DNA polymerase inhibitor) or hydroxyurea<sup>12,13</sup> (a ribonucleotide reductase inhibitor). APH or hydroxyurea generated similar levels of DNA damage in diploid cells, when compared with untreated tetraploid cells (Extended Data Fig. 2i). In addition to  $\gamma$ H2AX, we also observed a significant increase in the number of foci containing the double strand break repair factors FANCD2 and 53PB1<sup>14</sup> in the first interphase following WGD (Fig. 1i–l). Further, tetraploid cells showed an increased olive tail moment in alkaline comet assays, indicating single and double strand breaks (Extended Data Fig. 2j, k).

We next tested whether DNA damage is also generated in the subsequent cell cycles. A high proportion of tetraploid RPE-1 cells arrests after the first cell cycle in a LATS2–p53-dependent manner<sup>15</sup>. We thus analysed DNA damage levels in p53-depleted cells (Extended Data Fig. 2l). During the second and third interphases following tetraploidization, we observed a considerable decrease in DNA damage levels (Extended Data Fig. 2m–o). As most animal cells are normally organized in tissues with cell–cell adhesions, we tested the consequences of WGD in 3D cultures. Spheroids containing tetraploid cells displayed a higher  $\gamma$ H2AX index (Methods) compared with diploid cells (Extended Data Fig. 3a–d).

Collectively, our results show that a transition from a diploid to a tetraploid status after unscheduled WGD is accompanied by high levels of DNA damage within the first cell cycle.

## DNA replication-dependent DNA damage

We determined the cell cycle stage when the DNA damage occurs using the fluorescence ubiquitination cell cycle indicator (FUCCI). During G1, the number of  $\gamma$ H2AX foci was quite low and similar to that found in controls. As tetraploid cells entered S phase, we observed a slight increase in the number of foci, which increased substantially at the end of S phase (Fig. 2a, b, Extended Data Fig. 3e, f). These results were further confirmed by time-lapse imaging using H2B–GFP to visualize DNA and 53BP1–RFP (Extended Data Fig. 3g, h, Supplementary Videos 1, 2). To confirm that DNA damage in tetraploid cells appeared during S phase, we blocked cells at the G1/S

transition using high doses of inhibitors of CDK4/6 or CDK2 for 16 h (Extended Data Fig. 3*i,j*). We chose a 16-h period because this corresponds to the end of S phase in the cycling population (Fig. 2*a,b*) and thus enables us to distinguish whether DNA damage accumulates in a specific cell cycle phase or, alternatively, after a certain period of time. G1-arrested tetraploid cells showed low levels of DNA damage, whereas cells released in S phase exhibited high levels of DNA damage (Extended Data Fig. 3*i–o*). Of note, we observed a significant increase in the percentage of  $\gamma$ H2AX foci co-localizing with markers of active DNA replication sites visualized by proliferating cell nuclear antigen (PCNA) and EdU incorporation in tetraploid cells compared with diploid cells (31% versus 7%) (Extended Data Fig. 3*p,q*).

**Fig. 2: Genetic instability in tetraploid cells is generated during S phase in a DNA replication-dependent manner.**

figure 2



**a**, Top, DNA damage visualized using  $\gamma$ H2AX in RPE-1 tetraploid cells. Bottom, percentage of RPE-1 tetraploid cells in G1 or in S–G2. Data are mean  $>100$  interphase cells, 3 independent experiments. **b**, The number of  $\gamma$ H2AX foci per interphase cell in diploid (D) and tetraploid (T) RPE-1 cells. Data are mean  $\pm$  s.e.m.;  $>100$  interphase cells, 3 independent experiments. **c**, Percentage of RPE-1 tetraploid cells in G1 or in S–G2 and representative images showing DNA damage (anti- $\gamma$ H2AX) in tetraploid cells synchronized in G1 using 1  $\mu$ M palbociclib or released in S phase with or without

400 nM APH. Data are mean  $\pm$  s.e.m.; >100 interphase cells, 3 independent experiments. **d**, The number of  $\gamma$ H2AX foci per interphase cell in diploid and tetraploid RPE-1 cells from **c**. Data are mean  $\pm$  s.e.m.; >100 interphase cells, 3 independent experiments. **e**, Left, scheme for replication fork analysis. Right, immunofluorescence of DNA fibres obtained from diploid and tetraploid RPE-1 cells. **f, g**, Left, the replication fork speed in diploid and tetraploid RPE-1 (**f**) or BJ cells (**g**). Right, the CldU/IdU ratio in diploid and tetraploid RPE-1 (**f**) or BJ cells (**g**). Data are mean  $\pm$  s.e.m.; >330 replication forks (**f**), >295 replication forks (**g**). **i**, Genome-wide copy number plots for G2/M tetraploid RPE-1 or BJ cells induced by mitotic slippage. Each row represents a cell. Bottom right, workflow showing the method used to sort the cells. **b, d, f**, One-sided ANOVA test. **g**, Two-sided *t*-test. Scale bars, 10  $\mu$ m.

### Source data

By evaluating markers of DNA damage signalling and repair pathways we observed that the number of foci containing KU80 and XRCC1—proteins involved in non-homologous end joining<sup>16</sup>—remained low in tetraploid cells. By contrast, the number of foci containing the homologous recombination (HR) factor RAD51 was increased. Moreover, the percentage of RAD51 foci co-localizing with  $\gamma$ H2AX foci was significantly increased in tetraploid cells compared with diploid cells (14% versus 3%). Foci containing the replication stress markers replication protein A (RPA) and FANCD2 were also increased in number, and we observed a significant increase in their colocalization with  $\gamma$ H2AX foci in tetraploid cells compared with diploid cells (40% versus 14%) (Extended Data Fig. [4a–k](#)). Together, these results demonstrate that tetraploid cells experience high levels of DNA damage during S phase, indicated by markers of DNA damage and HR.

We hypothesized that DNA damage in tetraploid cells arises from errors occurring during DNA replication. To test this possibility, cells were arrested in G1 (Extended Data Fig. [3k](#)). We then released them in the presence of very low doses of APH or PHA-767491 (PHA; a Cdc7 inhibitor) to inhibit DNA replication (detected by absence of EdU) without generating DNA damage (Methods). This leads to inhibition of DNA replication while maintaining the biochemical activity typical of the S phase nucleus. DNA damage levels were markedly decreased in tetraploid cells treated with APH or PHA (Fig. [2c, d](#), Extended Data Fig. [5a–f](#)). Of note, in the few tetraploid cells that escaped DNA replication inhibition (revealed by high EdU incorporation) there was still a large number of  $\gamma$ H2AX foci (Extended Data Fig. [5g, h](#)). Together, these results establish that WGD generates DNA replication-dependent DNA damage. Deoxyribonucleoside triphosphate (dNTP) exhaustion leads to replication stress and genetic instability<sup>17</sup>. We tested whether supplying nucleosides rescued the DNA damage defects described above. This was however not the case in cells or in an *in vivo* model of polyploidy generation (Extended Data Figs. [5i, j](#), [10g](#)). These results

suggest that unscheduled WGD does not induce exhaustion of nucleoside levels as described in other oncogenic conditions<sup>17</sup>.

We characterized DNA replication using RPE-1 cell lines stably expressing PCNA chromobodies ([Supplementary Information](#), Methods). Quantitative 4D live imaging of DNA replication in diploid and tetraploid cells revealed marked decreases in the total number of PCNA foci and their volume and a similar effect on the number of EdU foci (Extended Data Fig. [6a–f](#)). This suggests a lack of scaling up with DNA content and fewer active replication sites in tetraploid cells. Time-lapse analysis of PCNA and fluorescence intensity was used as a readout of early and late S phase<sup>18</sup>, revealing a longer early S phase period in tetraploid cells ([Supplementary Information](#), Extended Data Fig. [6g–i](#), [Supplementary Videos 3, 4](#)). We next performed DNA combing, which enables visualization of replication fork behaviour in single DNA fibres<sup>13,19</sup>. Median fork speed and fork asymmetry (a readout of stalled or collapse forks) were increased in tetraploid cells (Fig. [2e–g](#), Extended Data Fig. [6j, k](#)). We attempted to analyse inter-origin distance (IOD), as the number of active regions can influence fork speed<sup>20</sup>. We noted a trend for increased IOD in tetraploid cells; however, it did not reach the threshold for significance (a possible explanation is provided in Methods).

To assess the type of karyotype generated in a single S phase after WGD, we used single-cell DNA sequencing (Methods, [Supplementary Information](#), Supplementary Methods). We identified over-duplicated chromosomes (more than 10) in addition to frequent over- and under-replicated regions (9n, 7n and 4n) in G2/M tetraploid cells (Fig. [2h](#), Extended Data Fig. [7a, b](#)). Both aneuploidy and heterogeneity scores and the proportion of the genome affected by aneuploidies were increased in G2/M tetraploid cells (Fig. [2h](#), Extended Data Fig. [7a–d](#), Methods). Our data establish that WGD generates abnormal karyotypes within a single S phase.

## Non-optimal S phase in tetraploid cells

Tetraploid cells would be expected to ‘scale up’ RNA and protein content by a factor of two. However, we found no evidence of such an increase in total RNA and protein content in newly born tetraploid cells using pyronin Y staining and quantitative phase imaging (Fig. [3a, b](#), Extended Data Fig. [8a–c](#)). We next tested the levels of key DNA replication factors. We developed protocols to sort tetraploids from diploids on the basis of FUCCI and DNA content from a common cell population (Fig. [3c](#), Extended Data Fig. [8d, e](#), Methods). The same number of cells was loaded for diploid and tetraploid conditions and total protein extracts and chromatin-bound extracts were probed by western blot. The chromatin-associated H2B variant, the cytoskeleton component actin and the membrane component β-catenin showed increases consistent with tetraploidization. By contrast, using H2B as a readout of DNA content, there was

no similar increase in G1 and S phase DNA replication factors in tetraploid cells (Fig. 3d–f, Extended Data Fig. 8f,g). We analysed the origin recognition complex 1<sup>21</sup> (ORC1), the minichromosome maintenance 2 helicase<sup>22</sup> (MCM2), Cdc10-dependent transcript 1 protein (CDT1)<sup>23</sup> and CDC6<sup>24</sup>. These proteins are key members of pre-replication complexes and are normally loaded in G1 during origin licensing. We also tested PCNA, CDC45<sup>25</sup> and treslin<sup>26</sup>, which are required for the initiation of DNA replication. We further probed the levels of E2F1, a transcription factor that activates the expression of S phase genes<sup>27,28,29</sup>. With the exception of treslin, the total levels of these proteins did not show the expected increase in tetraploid cells (Fig. 3e,f). Furthermore, levels of pre-replication complexes, treslin and CDC45 also did not increase in the chromatin-bound fractions from tetraploid cells (Fig. 3g,h, Extended Data Fig. 8h).

**Fig. 3: Key replication factors do not scale up in tetraploid cells.**

 figure 3

**a**, Tetraploid cells expressing FUCCI and the corresponding image under phase microscopy. **b**, The ratio of protein produced during G1 in diploid (D) and tetraploid (T) cells. Data are mean  $\pm$  s.e.m.; >50 G1 cells, 2 experiments. **c**, Schematic of fluorescence-activated cell sorting. **d**, Relative H2B levels in RPE-1 cells. Data are mean  $\pm$  s.e.m.; three experiments. **e, g**, Western blots of total protein extracts (e) or chromatin-bound extracts (g) obtained from RPE-1 cells. **f, h**, The protein levels from

total protein extracts in **e** (**f**) and chromatin-bound extracts in **g** (**h**). Data are mean  $\pm$  s.e.m.; three independent experiments. **i**, Stills from time-lapse videos of RPE-1 cells expressing FUCCI. **j**, Graph showing the duration of G1 in RPE-1 cells. Data are mean  $\pm$  s.e.m.; >35 interphase cells, 2 independent experiments. **k**, Graphs showing the time in G1 and the mass at birth of RPE-1 cells. More than 50 interphase cells, 2 independent experiments. **l**, **o**, Western blots of (**l**) or chromatin-bound extracts (**o**) obtained from RPE-1 cells with extended G1 duration. **m**, Relative H2B levels in RPE-1 cells with extended G1 duration. Data are mean  $\pm$  s.e.m.; four experiments. **n**, **p**, Protein concentration in total protein extracts from **l** (**n**) and chromatin-bound extracts from **o** (**p**). Data are mean  $\pm$  s.e.m.; three experiments. **q**, **r**, The number of  $\gamma$ H2AX foci in RPE-1 cells with G1 lengthening or G1 arrest using 160 nM or 1  $\mu$ M palbociclib and released in S phase. Tetraploidy induced by mitotic slippage (**q**) or endoreplication (**r**). Data are mean  $\pm$  s.e.m.; >100 interphase cells, 3 independent experiments. **e**, **g**, **l**, **o**, The same number of cells was loaded for each condition. **j**, **q**, **r**, One-sided ANOVA test. **d**, **f**, **h**, **m–o**, Two-sided *t*-test. **k**, Two-sided Pearson test. Scale bars, 50  $\mu$ m (**a**), 10  $\mu$ m (**i**).

### Source data

In normal proliferative cell cycles, growth occurring during G1 phase prepares cells for DNA replication, increasing the expression and accumulation of key S phase regulators<sup>28,30</sup>. We measured G1 duration in tetraploid cells and found only a slight increase compared with diploid cells (Fig. [3i,j](#), Extended Data Fig. [8i,j](#)). Further, although there was a significant correlation between cell mass and G1 duration in diploid cells, as described previously<sup>31</sup>, this was not the case in tetraploid cells (Fig. [3k](#)). We then tested whether G1 lengthening favoured error-free DNA replication in tetraploid cells. We delayed S phase entry using very low doses of inhibitors of CDK4/6 or CDK2 (Extended Data Fig. [9a–c](#), [Supplementary Information](#), Methods). In this condition, the levels of DNA replication factors from total cell or chromatin extracts scaled up with DNA content (comparing Fig. [3l–p](#) with Fig. [3e–h](#) and Extended Data Fig. [9k](#)). Further, the number and volume of active replication sites in S phase scaled up with DNA content in tetraploid cells and the dynamic behaviour of PCNA in tetraploid cells was similar to that in diploid cells (Extended Data Fig. [9d–h](#), [Supplementary Videos 5,6](#)). The time spent in S phase was not altered, but the ratio between early and late S phase in tetraploid cells was restored (Extended Data Fig. [9i,j](#)). In all cell lines, G1 lengthening was sufficient to reduce the number of  $\gamma$ H2AX, FANCD2 and 53BP1 foci in tetraploid S phase cells (Fig. [3q–r](#), Extended Data Fig. [9l–r](#)).

Our data show that tetraploid cells transition from G1 to S phase prematurely without undergoing scaling of global protein mass. They enter S phase with insufficient DNA replication factors, which can be compensated for by G1 lengthening.

## E2F1 rescues genetic instability in tetraploid cells

As the time spent in G1 does not prepare tetraploid cells for S phase, we reasoned that increased E2F1 levels might compensate for defects in G1 length scaling up. E2F1 is a transcription factor that promotes proliferation and cell cycle progression by regulating S phase and DNA replication factors<sup>28,29</sup>. We over-expressed E2F1 in diploid cells, enabling us to increase the expression of DNA replication proteins just before generating tetraploid cells. This was sufficient to rescue the levels of DNA damage in tetraploid cells (Fig. 4a–c, Extended Data Fig. 10a–c).

**Fig. 4: Increased E2F1 levels are sufficient to rescue genetic instability in both tetraploid cells and in unscheduled polyploid cells *in vivo*.**

figure 4



**a**, Top, workflow showing the method used to overexpress E2F1 (E2F1 OE). Bottom,  $\gamma$ H2AX immunofluorescence in cells overexpressing E2F1. **b, c**, Graphs showing the number of  $\gamma$ H2AX foci per interphase cell in diploid (D) and tetraploid (T) RPE-1 cells released in S phase with and without E2F1 overexpression. Tetraploidy induced by mitotic slippage (**b**) or endoreplication (**c**). Data are mean  $\pm$  s.e.m.; >100 interphase cells, three experiments. **d**, Experimental scheme to show the brain and the salivary glands of *Drosophila* larva. **e**, Representative images of salivary glands

from wild-type larvae and brain lobes of control or *sqh*-mutant larvae. **f**,  $\gamma$ H2Av index in salivary glands (SG) and in diploid (D) and polyploid (P) neural stem cells from the *Drosophila* larvae brain. NB, neuroblast. Data are mean  $\pm$  s.e.m.; >60 interphase cells, 3 experiments. **g**,  $\gamma$ H2Av in brain lobes of control or *sqh*-mutant larvae with or without E2F1 overexpression. **h**,  $\gamma$ H2Av index in neuroblasts with or without E2F1 overexpression. Data are mean  $\pm$  s.e.m.; >30 interphase cells, 3 experiments. **i**,  $\gamma$ H2Av in neuroblasts derived from *sqh*-mutant larvae with or without E2F1 overexpression. The yellow dotted lines indicate EdU-negative nuclei, the solid yellow line indicates EdU-positive nuclei. **j**,  $\gamma$ H2Av index in EdU-negative and EdU-positive nuclei with or without E2F1 overexpression. Data are mean  $\pm$  s.e.m.; >30 interphase cells, 3 experiments. **k**, Model in which a single S phase generates genetic instability in tetraploid cells. The white dotted lines indicate the nuclear (**a**) or cell area (**e, g, i**). **b**, **c**, **f**, **h**, **j**, One-sided ANOVA test. Scale bars, 10  $\mu$ m (**a, e** bottom right, **g** bottom), 20  $\mu$ m (**e** bottom middle, **i**), 50  $\mu$ m (**e** top, **e** bottom left, **g** top).

### Source data

A key prediction of our findings is that unscheduled polyploid *Drosophila* interphase neuroblasts<sup>32</sup> should also accumulate high levels of DNA damage in vivo. Indeed, the  $\gamma$ H2Av index (Methods) was higher in polyploid neuroblasts compared with diploid neuroblasts or programmed polyploid salivary gland cells, which normally accumulate very high ploidies<sup>33</sup> (Fig. 4d–f). We tested the effect of E2F1OE in polyploid neuroblasts and found that this was sufficient to decrease substantially DNA damage levels in vivo. Further, DNA damage was mainly restricted to EdU<sup>+</sup> nuclei (Fig. 4g–j, Extended Data Fig. 10d–f). Together, these data show that in vivo unscheduled polyploidy is a source of DNA damage and genetic instability in replicating cells, which can be inhibited by increased E2F1 levels.

As WGDs are quite frequent in human tumours, which have high levels of genetic instability<sup>1,2,34</sup>, our findings predict that these tumours must cope with increased DNA damage levels and therefore upregulate the DNA damage response pathway. We performed gene set enrichment analysis (GSEA) using cohorts of tetraploid and diploid lung, bladder and ovarian tumours<sup>35</sup>. This revealed an enrichment for DNA repair pathways in all tetraploid tumours when compared with diploid tumours (Extended Data Fig. 10h). These results suggest an increased requirement for the DNA damage response in tumours with WGD.

## Discussion

Here we analysed the initial defects following WGD and identified a very early window of high genetic instability that could promote acquisitions of multiple mutations, making it possible to bypass cell cycle controls while promoting survival of

tetraploid cells. Our results are consistent with a model in which tetraploid cells transit through the first cell cycle while lacking the capacity to support faithful replication of increased DNA content (Fig. 4k, [Supplementary Discussion](#)).

In non-physiological conditions, such as those studied here, newly born tetraploids might not sense the increase in DNA content and may therefore be unable to adapt G1 duration or protein content to replicate a 4n genome. Further research is needed to identify the molecular mechanisms that promote ploidy increase while maintaining genetic stability and cell homeostasis to understand how tetraploid cancers and tetraploids arising during evolution adapted to the new cellular state.

## Methods

### Cell culture

Cells were maintained at 37 °C in a 5% CO<sub>2</sub> atmosphere. hTERT RPE-1 cells (ATCC cat. no. CRL-4000, RRID:CVCL 4388) and HEK 293 cells (ATCC cat. no. CRL-1573, RRID:CVCL 0045) were grown in Dulbecco's modified medium (DMEM) F12 (11320-033 from Gibco) containing 10% fetal bovine serum (GE Healthcare), 100 U ml<sup>-1</sup> penicillin, 100 U ml<sup>-1</sup> streptomycin (15140-122 from Gibco). BJ cells (ATCC cat. no. CRL-4001, RRID:CVCL 6573) and HCT116 cells (ATCC cat. no. CCL-247, RRID:CVCL 0291) were grown in Dulbecco's modified medium + GlutaMAX (61965-026 from Gibco) containing 10% fetal bovine serum (GE Healthcare), 100 U ml<sup>-1</sup> penicillin, 100 U ml<sup>-1</sup> streptomycin (15140-122 from Gibco).

All cells were routinely checked for mycoplasma infection and are negative for mycoplasma infection. Identity and purity of the human cell lines used in this study were tested and confirmed using STR authentication.

### Generation of an RPE-1 PCNA<sup>chromo</sup> stable cell line

RPE-1 cells were transfected with 10 µg Cell Cycle-Chromobody plasmid (TagRFP) (from Chromotek) using JET PRIME kit (Polyplus Transfection, 114-07) according to the manufacturer's protocol. After 24 h, 500 µg ml<sup>-1</sup> G418 (4727878001 from Sigma Aldrich) was added to the cell culture medium and then a mixed population of clones expressing PCNA chromobodies were selected.

### Generation of an RPE-1 FUCCI or RPE-1 CCNB1<sup>AID</sup> FUCCI stable cell line

To produce lentiviral particles, HEK 293 cells were transfected with 4 µg pBOB-EF1-FastFUCCI-Puro (Addgene 86849) + 4 µg pMD2.G (Addgene 12259) + 4 µg psPAX2 (Addgene 12260) using a FuGENE HD Transfection Reagent (Promega E2311) in OptiMEM medium (ThermoFisher 51985034). Cells were incubated at 37 °C in a 5% CO<sub>2</sub> atmosphere for 16 h and then growth media were removed and replaced by 6 ml fresh OptiMEM. The following day, viral particles were isolated by filtering the medium containing them through a 0.45-µm filter (Sartorius Stedim Biotech 16537). Then, RPE-1 or RPE-1 CCNB1<sup>AID</sup> cells<sup>36</sup> were incubated with viral particles in the presence of 8 µg ml<sup>-1</sup> polybrene (Santa Cruz sc-134220) at 37 °C in a 5% CO<sub>2</sub> atmosphere for 24 h. RPE-1 GFP and RFP-positive cells were then collected using Sony SH800 FACS (BD FACSDiva Software Version 8.0.1). RPE-1 or RPE-1 CCNB1<sup>AID</sup> clones expressing FUCCI were selected and the cell lines were established from one single clone.

pBOB-EF1-FastFUCCI-Puro<sup>37</sup> was a gift from K. Brindle and D. Jodrell (Addgene 86849).

## **Generation of an RPE-1 GFP-53BP1 RFP-H2B stable cell line**

This cell line was obtained as described below. In brief, to produce lentiviral particles, HEK 293 cells were transfected with 4 µg pSMPUW-IRIS-Neo-H2BmRFP (Fachinetti laboratory) + 4 µg pMD2.G (Addgene 12259) + 4 µg psPAX2 (Addgene 12260). Then, RPE-1 cells were incubated with viral particles and RPE-1 RFP-positive cells were collected using Sony SH800 FACS (BD FACSDiva Software Version 8.0.1). RPE-1 clones expressing RFP-H2B were selected, and the cell line was established from one single clone.

Then, new lentiviral particles were produced by transfecting HEK 293 cells with 4 µg Apple-53BP1trunc (Addgene 69531) + 4 µg pMD2.G (Addgene 12259) + 4 µg psPAX2 (Addgene 12260). RPE-1 RFP-H2B cells were incubated with viral particles, and RPE-1 clones expressing both RFP-H2B and GFP-53BP1 were selected using flow cytometry (Sony SH800 FACS). The cell line was established from one single clone.

Apple-53BP1trunc was a gift from R. Weissleder<sup>38</sup> (Addgene).

## **Generation of an RPE-1 shp53 stable cell lines**

This cell line was obtained as described below. In brief, to produce lentiviral particles, HEK 293 cells were transfected with 4 µg short hairpin RNA (shRNA) p53-puromycin (Fachinetti laboratory) + 4 µg pMD2.G (Addgene 12259) + 4 µg psPAX2 (Addgene 12260). Then, RPE-1 cells were incubated with viral particles. After 24 h, 5 µg ml<sup>-1</sup>

puromycin (A1113803 from Gibco) was added to the cell culture medium and then a mixed population of clones expressing p53 shRNA was selected.

## Induction of tetraploidy in human cell lines

To induce mitotic slippage, cells were incubated with DMSO (D8418 from Sigma Aldrich) or with 50  $\mu$ M monastrol (S8439 from Selleckchem) + 1  $\mu$ M MPI-0479605 (S7488 from Selleckchem) for at least 2 h. Alternatively, CCNB1 depletion in RPE CCNB1<sup>AID</sup> cells was induced as described<sup>36</sup>. In brief, cells were treated with 2  $\mu$ g ml<sup>-1</sup> doxycycline (D3447 from Sigma Aldrich) + 3  $\mu$ M asunaprevir (S4935 from Selleckchem) for 2 h. Then, 500  $\mu$ M auxin (I5148 from Sigma Aldrich) was added to the cell culture medium for at least 4 h. In the figures, mitotic slippage was induced by the combination of monastrol + MPI-0479605 treatment except for the following figures: Figs. 2i, 3a–h, j–o, Extended Data Figs. 2a, b, 7a, d, 8d–h, 9k, in which mitotic slippage was induced by CCNB1 depletion.

To induce cytokinesis failure, cells were incubated with 10  $\mu$ M genistein (G6649 from Sigma Aldrich) for at least 2 h. Alternatively, cell were incubated with 0.75  $\mu$ M dihydrocytochalasin D (DCD; D1641 from Sigma-Aldrich) or with 5  $\mu$ M latrunculin (L5288 from Sigma-Aldrich) for 1 h. In the figures, cytokinesis failure was induced by genistein treatment except for the following figures: Extended Data Fig. 6j, k, in which cytokinesis failure was induced by DCD treatment and Extended Data Fig. 2c, d, in which cytokinesis failure was induced by latrunculin treatment.

To induce endoreplication, cells were incubated with 10  $\mu$ M SP600125 (S1460 from Selleckchem) for at least 2 h. Alternatively, CCNA2 depletion in RPE CCNA2<sup>AID</sup> cells was induced as described<sup>36</sup>. In brief, cells were treated with 2  $\mu$ g ml<sup>-1</sup> doxycycline (Sigma Aldrich D3447) for 2 h. Then, 500  $\mu$ M auxin (Sigma Aldrich I5148) + 3  $\mu$ M asunaprevir (Selleckchem S4935) was added to the cell culture medium for at least 4 h. In the figures, endoreplication was induced by SP600125 treatment except for Figs. 3q, 4c, Extended Data Figs. 2e, f, 3f, 5d, j, 8c, in which endoreplication was induced through CCNA2 depletion.

## Cell cycle synchronization and DNA replication inhibition

Cells were treated with 1  $\mu$ M palbociclib (Cdk4/6 inhibitor, Selleckchem S1579), or with 0.5  $\mu$ M abemaciclib (Cdk4/6 inhibitor, Selleckchem S5716) or with 1  $\mu$ M K03861(Cdk2 inhibitor, Selleckchem S8100) for 16 h to synchronize cells at G1/S transition, and were collected (indicated by ‘G1 arrest’ in the figures). Alternatively, cells were then washed five times with PBS and released in S phase for 10 h before being collected. To extend G1 duration cells were treated with 160 nM palbociclib or with 50 nM abemaciclib or with 400 nM K03861 for 16 h and were collected

(indicated by ‘G1 lengthening’ in the figures). Alternatively, cells were then washed 5 times in PBS and released in S phase for 10 h before being collected.

To inhibit DNA replication, cells were released in S phase in the presence of low doses of Aphidicolin (APH, A0781 from Sigma-Aldrich), a DNA replication polymerase inhibitor, or of PHA767491 (PZ0178 from Sigma-Aldrich), a Cdc7 inhibitor (indicated by ‘release in S phase + APH’ or ‘release in S phase + PHA’, respectively, in the figures). Doses were chosen to significantly decrease EdU incorporation without affecting the levels of DNA damage.

## Nucleoside supplementation

Cells were synchronized in G1 using 1  $\mu$ M palbociclib and then released in S phase (see ‘Cell cycle synchronization and DNA replication inhibition’) in the presence of nucleosides at the following concentrations: dC 7.3 mg l<sup>-1</sup> (Sigma Aldrich D0776); dG 8.5 mg l<sup>-1</sup> (Sigma Aldrich D0901); dU 7.3 mg l<sup>-1</sup> (Sigma Aldrich D5412); dA 8 mg l<sup>-1</sup> (Sigma Aldrich D8668) and dT 2.4 mg l<sup>-1</sup> (Sigma Aldrich T1895) (+ in the figures) or dC 14.6 mg l<sup>-1</sup>; dG 17 mg l<sup>-1</sup>; dU 14.6 mg l<sup>-1</sup>; dA 16 mg l<sup>-1</sup> and dT 4.8 mg l<sup>-1</sup> (++ in the figures).

## Treatments

The drugs were used at the following concentrations: Auxin (Sigma I5148), 500  $\mu$ M; doxycycline (Sigma D3447), 2  $\mu$ g ml<sup>-1</sup>; asunaprevir (Selleckchem S4935), 3  $\mu$ M; monastrol (Selleckchem S8439), 50  $\mu$ M; MPI-0479605 (Selleckchem S7488), 1  $\mu$ M; genistein (Sigma G6649), 10  $\mu$ M; SP600125 (Selleckchem S1460), 10  $\mu$ M; abemaciclib (Selleckchem S5716), 50 nM or 0.5  $\mu$ M; K03861 (Selleckchem S8100), 400 nM or 1  $\mu$ M; palbociclib (Selleckchem S1579), 120 nM or 1  $\mu$ M; aphidicolin (Sigma A0781), 0.4  $\mu$ M or 1  $\mu$ M; hydroxyurea (Selleckchem S1896), 2 mM; PHA767491 (Sigma PZ0178), 1  $\mu$ M; RO3306 (Calbiochem 217699), 10  $\mu$ M; dihydrocytochalasin D (Sigma D1641), 0.75  $\mu$ M; latrunculin B (Sigma L5288), 5  $\mu$ M; 5'-chloro-2'-deoxyuridine (CIdU) (Sigma C6891), 100  $\mu$ M; 5'-ido-2'-deoxyuridine (IdU) (Sigma I7125), 100  $\mu$ M.

## Fly husbandry and fly stocks

Flies were raised on cornmeal medium (0.75% agar, 3.5% organic wheat flour, 5.0% yeast, 5.5% sugar, 2.5% nipagin, 1.0% penicillin-streptomycin and 0.4% propionic acid). Fly stocks were maintained at 18 °C. Crosses were carried out in plastic vials and maintained at 25 °C. Stocks were maintained using balancer inverted chromosomes to prevent recombination. Stocks used in this study: *sqh*<sup>1,39</sup>, *pavarotti*

RNAi (Pav<sup>RNAi</sup>) (Bloomington *Drosophila* Stock Center BL#42573)<sup>32</sup>, UAS-E2F1 (FlyORF F001065) and UAS-Rb (Bloomington *Drosophila* Stock Center BL#50746).

In all experiments, larvae were staged to obtain comparable stages of development. Egg collection was performed at 25 °C for 24 h. After development at 25 °C, third instar larvae were used for dissection.

## Preparation and imaging of human cells

Cells were plated on cover slips in 12-well plates and treated with the indicated drugs. To label cells, they were fixed using 4% of paraformaldehyde (Electron Microscopy Sciences 15710) + Triton X-100 (2000-C from Euromedex) 0.1% in PBS (20 min at 4 °C). Then, cells were washed three times using PBS-T (PBS + 0.1% Triton X-100 + 0.02% Sodium Azide) and incubated with PBS-T + BSA (Euromedex 04-100-812-C) 1% for 30 min at room temperature. After 3 washes with PBS-T + BSA, primary and secondary antibodies were incubated in PBS-T + BSA 1% for 1 h and 30 min at room temperature, respectively. After 2 washes with PBS, cells were incubated with 3 µg ml<sup>-1</sup> DAPI (Sigma Aldrich D8417) for 15 min at room temperature. After two washes with PBS, slides were mounted using 1.25% *n*-propyl gallate (Sigma P3130), 75% glycerol (bidistilled, 99.5%, VWR 24388-295), 23.75% H<sub>2</sub>O.

Images were acquired on an upright widefield microscope (DM6B, Leica Systems, Germany) equipped with a motorized *xy* stage and a 40× objective (HCX PL APO 40×/1.40–0.70 Oil from Leica). Acquisitions were performed using Metamorph 7.10.1 software (Molecular Devices) and a sCMOS camera (Flash 4V2, Hamamatsu). Stacks of conventional fluorescence images were collected automatically at a *z*-distance of 0.5 µm (Metamorph 7.10.1 software; Molecular Devices, SCR 002368). Images are presented as maximum intensity projections generated with ImageJ software (SCR 002285).

## Whole-mount tissue preparation and imaging of *Drosophila* larval brains

Brains or salivary glands from third instar larvae were dissected in PBS and fixed for 30 min in 4% paraformaldehyde in PBS. They were washed 3 times in PBST 0.3% (PBS, 0.3% Triton X-100 (Sigma T9284), 10 min for each wash) and incubated for several hours in agitation at room temperature and overnight at 4 °C with primary antibodies at the appropriate dilution in PBST 0.3%. Tissues were washed three times in PBST 0.3% (10 min for each wash) and incubated overnight at 4 °C with secondary antibodies diluted in PBST 0.3%. Brains and salivary glands were then washed 2 times in PBST 0.3% (30 min for each wash), rinsed in PBS and incubated with 3 µg ml<sup>-1</sup> DAPI (4',6-diamidino-2-phenylindole; Sigma Aldrich D8417) at room temperature for

30 min. Brains and salivary glands were then washed in PBST 0.3% at room temperature for 30 min and mounted on mounting media. A standard mounting medium was prepared with 1.25% *n*-propyl gallate (Sigma P3130), 75% glycerol (bidistilled, 99.5%, VWR 24388-295), 23.75% H<sub>2</sub>O.

Images were acquired on a spinning disk microscope (Gataca Systems). Based on a CSU-W1 (Yokogawa), the spinning head was mounted on an inverted Eclipse Ti2 microscope equipped with a motorized *xy* stage (Nikon). Images were acquired through a 40× NA 1.3 oil objective with a sCMOS camera (Prime95B, Photometrics). Optical sectioning was achieved using a piezo stage (Nano-z series, Mad City Lab). The Gataca Systems' laser bench was equipped with 405, 491 and 561 nm laser diodes, delivering 150 mW each, coupled to the spinning disk head through a single mode fibre. Multi-dimensional acquisitions were performed using Metamorph 7.10.1 software (Molecular Devices). Stacks of conventional fluorescence images were collected automatically at a *z*-distance of 1.5 μm (Metamorph 7.10.1 software; Molecular Devices SCR 002368). Images are presented as maximum intensity projections generated with ImageJ software (SCR 002285).

Primary and secondary antibodies were used at the following concentrations: guinea pig anti-CEP192 antibody<sup>40</sup> (1:500; R.B. laboratory), rabbit anti-β catenin (1:250; Sigma-Aldrich C2206, RRID AB 476831), mouse anti-γH2A.X phospho S139 (1:1,000; Abcam ab22551, RRID AB 447150), mouse anti-XRCC1 (1:500; Abcam ab1838, RRID AB 302636), rabbit anti-Rad51 (1:500; Abcam ab133534, RRID AB 2722613), mouse anti-KU80 (1:200; ThermoFisher MA5-12933, RRID AB 10983840), rabbit anti-FANCD2 (1:150; Novusbio NB100-182SS, RRID AB 1108397), mouse anti-53BP1 (1:250; Millipore MAB3802, RRID AB 2206767), rabbit anti-γH2Av (1:500; Rockland600-401-914, RRID AB 11183655), Alexa Fluor 647 Phalloidin (1:250; ThermoFisher Scientific A22287, RRID AB 2620155), goat anti-rabbit IgG (H+L) Highly Cross-Adsorbed Secondary Antibody, Alexa Fluor 647 (1:250; ThermoFisher A21245, RRID AB 2535813), goat anti-guinea pig IgG (H+L) Highly Cross-Adsorbed Secondary Antibody, Alexa Fluor 488 (1:250; ThermoFisher A11073, RRID AB 253411), goat anti-mouse IgG (H+L) Cross-Adsorbed Secondary Antibody, Alexa Fluor 546 (1:250; ThermoFisher A11003, RRID AB 2534071), goat anti-rabbit IgG (H+L) Highly Cross-Adsorbed Secondary Antibody, Alexa Fluor 546 (1:250; Thermo Fisher Scientific A-11035, RRID AB 2534093).

## Quantitative analysis of DNA damage

### *Drosophila* neuroblasts and 3D spheroids

Quantitative analysis of DNA damage was carried out as previously described<sup>32</sup>. In brief, DNA damage was assessed in *Drosophila* using a γH2Av primary antibody and

in 3D spheroids with a  $\gamma$ H2AX antibody, and detected with an Alexa Fluor secondary antibody. Confocal volumes were obtained with optical sections at 1.5- $\mu$ m intervals. Image analysis was performed using Fiji and a custom plugin developed by QUANTACELL. After manual segmentation of the nuclei, a thresholding operation was used to determine the percentage of  $\gamma$ H2Av- or  $\gamma$ H2AX-positive pixels (coverage) and their average intensity in a single projection. Coverage and intensity were multiplied to obtain the  $\gamma$ H2Av or  $\gamma$ H2AX index. The threshold used to detect and quantify the  $\gamma$ H2Av index in polyploid neuroblasts does not detect any damage in salivary glands. However, it is important to mention that in a fraction of these cells,  $\gamma$ H2Av dots (small and of low fluorescence intensity) can be occasionally seen.

## 2D human cell lines

For DNA damage quantification, the signals obtained in cultured cells were different from the signals found in *Drosophila* neuroblasts. To asses DNA damage in human cells, we used an ImageJ software-based plugin developed by QUANTACELL, where  $\gamma$ H2AX signals were measured using z-projection stacks after thresholding. Nuclear size, DAPI intensity, the number of  $\gamma$ H2AX foci,  $\gamma$ H2AX fluorescence intensity and the percentage of nuclear coverage by  $\gamma$ H2AX signal were obtained for each nucleus.

## Time-lapse microscopy

Cells were plated on a dish (627870 from Dutscher) and treated with the indicated drugs. Images were acquired on a spinning disc microscope (Gataca Systems). Based on a CSU-W1 (Yokogawa), the spinning head was mounted on an inverted Eclipse Ti2 microscope equipped with a motorized xy stage (Nikon). Images were acquired through a 40 $\times$  NA 1.3 oil objective with a sCMOS camera (Prime95B, Photometrics). Optical sectioning was achieved using a piezo stage (Nano-z series, Mad City Lab). Gataca Systems' laser bench was equipped with 405-, 491- and 561-nm laser diodes, delivering 150 mW each, coupled to the spinning disk head through a single mode fibre. Laser power was chosen to obtain the best ratio of signal/background while avoiding phototoxicity. Multi-dimensional acquisitions were performed using Metamorph 7.10.1 software (Molecular Devices). Stacks of conventional fluorescence images were collected automatically at a z-distance of 0.5  $\mu$ m (Metamorph 7.10.1 software; Molecular Devices, RRID SCR 002368). Images are presented as maximum intensity projections generated with ImageJ software (RRID SCR 002285), from stacks deconvolved with an extension of Metamorph 7.10.1 software.

## 3D cultures

### Mitotic slippage on 3D cultures

To generate spheroids, 500 cells per well were seeded into 96 ultra-low-attachment well plates (Corning7007) in presence of DMSO (Sigma Aldrich D8418) or with 50  $\mu$ M monastrol (Selleckchem S8439) and 1  $\mu$ M MPI-0479605 (Selleckchem S7488). Plates were spin down at 200g for 3 min, to allow spheroid formation, and incubated for 24 h at 37 °C.

### **Immunostaining**

Spheroids were collected and washed quickly with PBS before fixation using 4% paraformaldehyde (Electron Microscopy Sciences 15710) in PBS for 40 min. Then, spheroids were permeabilized for 5 min using Triton X-100 (Euromedex 2000-C) 0.3% in PBS and blocked for 30 min using blocking buffer (PBS + 0.3% Triton X-100 + 0.02% sodium azide + 3% BSA). Aggregates were incubated with primary antibodies diluted into blocking buffer overnight. After 3 washes using blocking buffer, spheroids were incubated with secondary antibodies in blocking buffer for 3 h. Cells were then washed several times for 2 h in blocking buffer and mounted on glass with EverBrite (Biotium). For primary and secondary antibodies see ‘Immunofluorescence microscopy and antibodies’.

### **Imaging and DNA damage analysis**

Spheroids were imaged using an inverted scanning laser confocal (Nikon A1RHD25) equipped with a 100 $\times$  CFI Plan Apo Lambda S Sil objective (NA 1.35). z-stacks were acquired every 0.3  $\mu$ m. Diploid and tetraploid cells were distinguished using cell and nuclear size and centrosome number. Then, quantitative analysis of DNA damage was carried out (see ‘Quantitative analysis of DNA damage’).

### **EdU staining**

EdU incorporation into DNA was visualized with the Click-it EdU imaging kit (Life Technologies C10338), according to the manufacturer’s instructions. For human cell lines, EdU was used at a concentration of 1  $\mu$ M (Extended Data Figs. 6e, 9h) or 10  $\mu$ M (Extended Data Fig. 5g, h) for the indicated time. Cells were incubated with the Click-it reaction cocktail for 15 min. EdU incorporation in polyploid neuroblasts was done as previously described<sup>32</sup> with a pulse of 2 h before fixation.

### **Comet assay**

Comet assays were performed using Single Cell Gel Electrophoresis Assay kit (4250-050-ES from Trevigen) according to the manufacturer’s instructions. Comets were then imaged using an inverted Eclipse Ti-E Nikon videomicroscope equipped with a

40× CFI Plan Fluor objective. Images were analysed with OpenComet plugin on Fiji. Based on the comet DNA content of DMSO treated cells, a manual threshold was applied to identify diploid from tetraploid cells. The same threshold was applied on the cells treated for mitotic slippage.

## FACS of diploid and tetraploid cells

A mix of diploid and tetraploid cells (see ‘Induction of tetraploidy in human cell lines’) were incubated with 2 µg ml<sup>-1</sup> Hoescht 33342 (Sigma Aldrich 94403) for 1 h at 37 °C, 5% CO<sub>2</sub>. Then, a single cell suspension was generated. Cells were washed using PBS, the supernatant was removed and cells were resuspended in a cold cell culture medium at 1 × 10<sup>7</sup> cell per ml and kept at 4 °C during all the experiments. Fluorescence-activated cell sorting (FACS) was performed using Sony SH800 FACS (BD FACSDiva Software Version 8.0.1). Compensation was performed using the appropriate negative control samples. Experimental samples were then recorded and sorted using gating tools to select the populations of interest. RFP<sup>+</sup>GFP<sup>-</sup> cells (G1 cells) were first selected. Then, in this population, DNA content was used to segregate diploid (2n) and tetraploid (4n) G1 cells (Extended Data Fig. 8d). Once gates have been determined, the same number of diploid and tetraploid G1 cells were sorted into external collection tubes. The number of cells was then checked using a cell counter and the same number of diploid and tetraploid cells were collected for western blot analysis. In parallel, post-sort analysis was performed to determine the purity of the sorted populations (Extended Data Fig. 8e).

## Cell cycle analysis and measure of RNA levels by flow cytometry

Cells were detached by treatment with Accutase (Sigma), immediately washed in PBS, fixed in 2 ml 70% ethanol and stored at -20 °C overnight. They were then washed in PBS and staining buffer (BD Pharmingen 554656).

For cell cycle analysis, DNA content was visualized by incubating the cells with 2 µg ml<sup>-1</sup> Hoescht 33342 (Sigma Aldrich 94403) in staining buffer for 15 min at room temperature. Alternatively, to measure RNA levels, cells were incubated with 2 µg ml<sup>-1</sup> Hoescht 33342 + pyronin 4 µg ml<sup>-1</sup> (Santa Cruz sc-203755A) in a staining buffer for 20 min at room temperature. Flow cytometry analysis was done using LSRII (BD Biosciences), by analysing 10,000 cells per condition. Data were then analysed with FlowJo 10.6.0 software (Tree Star).

## E2F1 overexpression

RPE-1 cells were transfected using 0.25 µg pCMVHA E2F1 (Addgene 24225) with a JET PRIME kit (Polyplus Transfection 114-07) according to the manufacturer's protocol. Five hours later, cells were incubated with DMSO (D8418 from Sigma Aldrich) or with 50 µM monastrol (Selleckchem S8439) + 1 µM MPI-0479605 (Selleckchem S7488) to generate tetraploid cells. After 2 h, DMSO or 1 µM palbociclib (Selleckchem S1579) were added to the cell culture medium for 16 h. Cells were then fixed in G1 ( $T_0$ ) or washed five times using PBS and released in S phase and fixed after 10 h ( $T_{10}$ ). The immunofluorescence protocol is described in the corresponding section.

pCMVHA E2F1 was a gift from K. Helin<sup>41</sup> (Addgene plasmid 24225).

## Western blot

For a whole-cell extract, cells were lysed in 8 M urea, 50 mM Tris HCl, pH 7.5 and 150 mM β-mercaptoethanol (Bio-Rad 161-0710), sonicated and heated at 95 °C for 10 min. For chromatin-bound fractions, cells were prepared using the Subcellular Protein Fractionation Kit for Cultured Cells (ThermoFisher Scientific 78840), according to the manufacturer's instructions. Then, samples (equivalent of  $2 \times 10^5$  cells) were subjected to electrophoresis in NuPAGE Novex 4–12% Bis-Tris pre-cast gels (Life Technologies NP0321). The same number of cells (see 'FACS sorting of diploid and tetraploid cells') were loaded for diploid and tetraploid conditions, allowing us to compare one diploid cell with one tetraploid cell. Protein fractions from the gel were electrophoretically transferred to PVDF membranes (PVDF transfer membrane; GE Healthcare RPN303F). After 1 h saturation in PBS containing 5% dry non-fat milk and 0.5% Tween 20, the membranes were incubated for 1 h with a primary antibody diluted in PBS containing 5% dry non-fat milk and 0.5% Tween 20. After three 10-min washes with PBS containing 0.5% Tween 20, the membranes were incubated for 45 min with a 1:2,500 dilution of peroxidase-conjugated antibody. Membranes were then washed three times with PBS containing 0.5% Tween 20, and the reaction was developed according to the manufacturer's specifications using ECL reagent (SuperSignal West Pico Chemiluminescent Substrate; Thermo Scientific 34080).

The background-adjusted volume intensity was calculated and normalized using a H2B signal (H2B was used as a readout of DNA content) for each protein, using Image Lab software version 6.0.1, Bio-Rad Laboratories. All the original uncropped blots (gel source data) are presented in Supplementary Fig. 1.

Primary and secondary antibodies were used at the following concentrations. Mouse anti-α-tubulin (1:5,000; Sigma T9026, RRID AB 477593), mouse anti-CDC45 (1:100; Santa Cruz Biotechnology sc-55569, RRID AB 831146), rabbit anti-PCNA (1:500; Santa Cruz sc56, RRID AB 628110), rabbit anti-actin (1:2,000; Sigma-Aldrich A5060,

RRID AB 476738), mouse anti-H2B (1:1,000; Santa Cruz Biotechnology sc-515808), mouse anti-ORC1 (1:100; Santa Cruz Biotechnology sc-398734), mouse anti-MCM2 (1:500; BD Biosciences 610701, RRID AB 398024), mouse anti-E2F1 (1:2,000; Santa Cruz sc251, RRID AB 627476), mouse anti-CDC6 (1:500; Santa Cruz sc-9964, RRID AB 627236), rabbit anti-CDT1 (1:500; Cell Signaling 8064S, RRID AB 10896851), rabbit anti-treslin (1:500; Betyl A303-472A, RRID AB 10953949), goat anti-rabbit IgG (H+L) Cross-Adsorbed Secondary Antibody, HRP (1:2,500; ThermoFisher G21234, RRID AB 2536530), Peroxidase AffiniPure goat anti-mouse IgG (H+L) (1:2500; Jackson ImmunoResearch 115-035-003, RRID AB 10015289).

### 3D reconstruction and analysis

3D videos (see ‘Time-lapse microscopy’) were imported into Imaris software v.9.6.0 (Bitplane, RRID SCR 007370). For chosen cells, the module ‘Spot tracking’ of Imaris v.9.6.0 was used to detect the foci, as spots of diameter 0.5 µm in the *xy*-direction and 1 µm in *z*-direction (modelling PSF elongation). Because the volume of the foci changes in time, the option ‘Enable growing regions’ was used. In each video, the threshold was chosen on the brightest frame (to detect a maximum of the correct spots) and then applied to the whole video. For each cell, at each time point, the number of spots and volumes were recorded. To determine DNA replication timing, we quantified the signal of PCNA fluorescence intensity in the nucleus. This replication timing was characterized independently of any particular behaviour of PCNA. As soon as PCNA fluorescence intensity was detected in the nucleus,  $t = 0$  (beginning of S phase) was defined, and when PCNA fluorescence intensity was not detected anymore the last time point was defined (end of S phase). For each condition, at least ten cells (Supplementary Data 1) were studied and the statistics from Imaris v.9.6.0 were averaged at each time point using a MATLAB script.

### Molecular combing

Tetraploid HCT116 were generated by cytokinesis inhibition using 0.75 µM dihydrocytochalasin D (DCD, inhibitor of actin polymerization, Sigma-Aldrich D1641) for 18 h overnight. Afterwards, the cells were washed 3 times with PBS and cultured in DMEM supplemented with 10% FBS and 1% penicillin-streptomycin for additional 20 h. Tetraploid RPE-1 and BJ cells were generated by mitotic slippage or endoreplication (see ‘Induction of tetraploidy in human cell lines’). Then, the cells were washed three times with PBS and cultured in DMEM supplemented with 10% FBS and 1% penicillin-streptomycin for an additional 20 h. For each method, we determined that the proportion of tetraploid cells in the treated population is about 40–60%. Due to the presence of diploid cells in the treated population, the consequences of tetraploidization on replication fork speed, fork asymmetry and IOD are most probably underestimated.

Diploid controls and the tetraploid-enriched population were then pulse-labelled with 0.1 mM CldU and 0.1 mM IdU for 30 min and 100,000–300,000 cells per condition were collected for further analysis. The DNA was extracted from cells and prepped following the manufacturer’s instructions using the FiberPrep DNA Extraction Kit (Genomic Vision). Subsequently, the prepped DNA was stretched onto coated glass coverslips (CombiCoverslips, Genomic Vision) by using the FiberComb Molecular Combing System (Genomic Vision). The labelling was performed with antibodies against ssDNA, IdU and CldU using the Replication Combing Assay (RCA) (Genomic Vision). The imaging of the prepared cover slips was carried out by Genomic Vision and analysed using the FiberStudio 2.0.1 Analysis Software by Genomic Vision. Replication speed was determined by measuring the combined length of the CldU and IdU tracks. Fork asymmetry was determined by measuring symmetry of the CldU and IdU incorporation by the forks (the length of the first track (CldU) is compared to the length of the second track (IdU)). IOD was determined by measuring distance between two origins on the same fibres.

Antibodies were used at the following concentrations. Rabbit anti-ssDNA (1:5; IBL International 18731, RRID AB 494649), rat anti-CldU (1:10; Abcam Ab6326, RRID AB 2313786), mouse anti-IdU (1:10; BD Biosciences 555627, RRID AB 10015222), mouse Alexa Fluor 647 donkey (1:25; Biozol JIM-715-605-151), rat Alexa Fluor 594 donkey (1:25; Biozol JIM-712-585-153), rabbit Brilliant Violet 480 donkey (1:25; Jackson Immuno Research 711-685-152, RRID AB 2651109).

## Quantitative phase imaging and measurements

Cells were plated on glass-bottom dishes coated with 50 µg ml<sup>-1</sup> fibronectin for 1 h and rinsed, and trypsinized cells were plated at a concentration of 1.5 × 10<sup>6</sup> cells per ml. The cells used for the experiments were seeded in T-25 dishes at a concentration of 0.7 × 10<sup>6</sup> cells per ml 2 days before the actual experiment. On the day of the experiment, the cells were detached with EDTA (versene), and plated at a concentration of 1.5 × 10<sup>6</sup> cells per ml. For inducing tetraploidy, cells were treated with 2 µg ml<sup>-1</sup> doxycycline (Sigma Aldrich D3447) for 2 h. Then, 500 µM auxin (Sigma Aldrich I5148) + 3 µM asunaprevir (Selleckchem S4935) was added to the cell culture medium for at least 4 h. The cells were then imaged for 35 h every 20 min to track them throughout their cell cycle.

The cell cycle state was indicated by the FUCCI system; G1 cells express Cdt1–RFP while S/G2 cells express geminin–GFP and mitosis was indicated by the nuclear envelope break down with geminin being present through the cells<sup>42</sup>. To quantify the fluorescence of geminin in the nucleus, first a background subtraction was performed on the images. A region of interest (ROI) was used to define an area containing the background fluorescence in the image. An average value of the ROI was then

subtracted from all the frames. Subsequently, a ROI was drawn as close as possible to the cell, and then the mean gray value was measured across all the frames. This helped identify the frames of birth and G1/S transition during the cell cycle.

A detailed protocol for the mass measurement with phasics camera is available in refs.<sup>43,44</sup>. Images were acquired by a Phasics camera every 20 min for 35 h for the duration of the experiment. To obtain the reference image, 32 empty fields were acquired on the dish and a median image was calculated. This reference image was subtracted from the interferograms (images acquired by phasics) by custom written MATLAB scripts to measure the optical path difference. They were then processed to calculate the phase, intensity and phase cleaned images (the background set to 1,000 and the field cropped to remove edges). Background normalization was performed using a gridfit method, and a watershed algorithm was used to separate cells which came in contact with each other. Mass was calculated by integrating the intensity of the whole cell.

## Sequencing and AneuFinder analysis

A mixed population of diploid and tetraploid RPE-1 CCNB1<sup>AID</sup> FUCCI cells were synchronized in G1 using 1 µM palbociclib (Selleckchem S1579) for 16 h or released in S phase for 20 h in the presence of 10 µM RO3306 (Calbiochem 217699) in order to block cells in the subsequent G2/M. G1 and G2/M diploid and tetraploid cells were then isolated using cell sorting (see ‘FACS sorting of diploid and tetraploid cells’) and collected in a 96-well plate.

Sequencing was performed using a NextSeq 500 (Illumina; up to 77 cycles; single end). The generated data were subsequently demultiplexed using sample-specific barcodes and changed into fastq files using bcl2fastq (Illumina; version 1.8.4). Reads were afterwards aligned to the human reference genome (GRCh38/hg38) using Bowtie2 (version 2.2.4; ref. <sup>45</sup>). Duplicate reads were marked with BamUtil (version 1.0.3; ref. <sup>46</sup>). The aligned read data (bam files) were analysed with the copy number calling algorithm AneuFinder<sup>47</sup> (<https://github.com/ataudt/aneufinder>). Following GC correction and blacklisting of artefact-prone regions (extreme low or high coverage in control samples), libraries were analysed using the dnacopy and ed divisive copy number calling algorithms with variable width bins (average bin size = 1 Mb; step size = 500 kb). The G1 samples were analysed with an euploid reference<sup>48</sup>. The G1 samples were used as a reference for the analysis of the G2/M samples (G1 diploid for G2/M diploid and G1 polyploid for G2/M polyploid). Aneuploid libraries were not used as a reference and blacklists were constructed using the example from Bioconductor as a guideline. The RPE-1 diploid G1 sample (2n) was analysed with the standard version of AneuFinder (from Bioconductor) while the other samples were analysed with the developer version of AneuFinder (from GitHub; 4n and 8n samples).

The ground ploidy for these samples was constrained between 3.5 and 4.5 (4n samples) or between 7.5 and 8.5 (8n samples; parameters: min.ground.ploidy and max.ground.ploidy). Results were afterwards curated by requiring a minimum concordance of 95 % (2n sample) or 90% (4n and 8n samples) between the results of the two algorithms. Libraries with on average less than 10 reads per chromosome copy of each bin (2-somy: 20 reads, 3-somy: 30 reads, etc.) were discarded. This minimum number of reads comes down to roughly 60,000 for a diploid genome in G1 phase (2n) up to 240,000 for a polyploid genome in G2/M phase (8n). Analysis of the BJ samples showed aberrations (wavy patterns) that resulted in wrongly called segments with a copy number which is either one higher or one lower than the expected state (when euploid). The means of the read counts (read counts of the bins) of these states were too close to the mean of the expected state (for example, mean 5-somy too close to mean 4-somy; 4n sample; Supplementary Methods 1). When more than 1 % of the genome was classified as such (for example, more than 1 % 5-somy), a non-rounded version of the copy number of the state was calculated using the mean of the expected state (ploidy of euploid sample) as a reference:

Non-rounded copy number.state = Mean.state/(mean.expected.state/copy.number.expected.state)

Example 5-somy (4n sample):

Non-rounded copy number.5-somy = Mean.5-somy/(Mean.4-somy/4)

This was done to quantify the distance between the two states. The values are typically between -0.5 and +0.5 of the state under consideration (for example, 5-somy; between 4.5 and 5.5), which will result in a rounded value equal to the state. The libraries with aberrations have typically a deviation of 0.25 and more from the expected value (Supplementary Methods 1). Libraries that showed a deviation of more than 0.25 were therefore discarded (For 5-somy; a value lower than 4.75 or higher than 5.25). By applying this cut-off, we eliminated libraries that clearly showed this aberration (Supplementary Methods 1) while preserving true aneuploid libraries (Supplementary Methods 1). This specific method was only used for the BJ samples.

## GSEA with TCGA PanCancer data

GSEA was performed using GSEA software v.4.2.1<sup>49,50</sup>. The normalized mRNA expression (Illumina HiSeq\_RNASeqV2, RSEM) from pan cancer studies were downloaded from <https://www.cbiportal.org/>; detailed information about RNA sequencing experiment and tools used can be found at the NCI's Genomic Data Commons (GDC) portal <https://gdc.cancer.gov>. The ploidy status for bladder urothelial carcinoma (156 near-diploid and 200 near-tetraploid samples), Lung adenocarcinoma (205 near-diploid and 240 near-tetraploid samples), and ovarian

serous cystadenocarcinoma (116 near-diploid and 130 near-tetraploid samples) were extracted from<sup>35</sup>. In addition to ranked list of genes and ploidy status, we use gene sets derived from the GO Biological Process ontology to assess significant pathway enrichment between near -diploid and near tetraploid tumors in GSEA tool. GSEA is a computational method that determines whether a defined set of genes shows statistically significant concordant differences between two biological states (for example, two distinct phenotypes), using the algorithm based on the calculation of an enrichment score (ES), the estimation of significance level of ES (nominal *P* value) and adjustment for multiple hypothesis testing (ES normalization and FDR calculation)<sup>49</sup>.

## Quantifications

Image analysis and quantifications were performed using Image J software V2.1.0/1.53c, <https://imagej.net/software/fiji/downloads>. To quantify the colocalizations between two signals (Extended Data Figs. 3i, m, 4g, j) we calculated the Manders coefficient using the JACOP plugin with Image J V2.1.0/1.53c software. We determined that the colocalizations between  $\gamma$ H2AX signal and EdU, FANCD2 or RAD51 signals are not random using an home-made based Costes randomization on nuclear area with Image J software. 1000 randomizations of the pixel positions were performed for each condition (Supplementary data 2). 3D videos (Extended Data Figs. 3c, 6c, 9c, d) were corrected using the 3D correct drift plugin with Image J V2.1.0/1.53c software to keep the cell of interest at the centre of the region of interest. The nuclear area and DAPI intensity were measured using the wand tool with Image J V2.1.0/1.53c software. For the figures, images were processed on Image J V2.1.0/1.53c software, and mounted using Affinity Designer (<https://affinity.serif.com/fr/designer/>).

## Statistics and reproducibility

At least two (*n*) independent experiments were carried out to generate each dataset, and the statistical significance of differences was calculated using GraphPad Prism (RRID SCR 002798) version 7.00 for Mac (GraphPad Software). The statistical test used for each experiment is indicated in the figure legends. Each representative image (Figs. 2a, c, 3a, e, g, k, n, 4a, Extended Data Figs. 2a, c, e, l, 3a, g, 4c, 5g, 6c, 9c, d, 10a) originates from a dataset composed of at least two (*n*) independent experiments.

## Reporting summary

Further information on research design is available in the [Nature Research Reporting Summary](#) linked to this paper.

## Data availability

Source data are available at <https://doi.org/10.6084/m9.figshare.19137323.v1>. [Source data](#) are provided with this paper.

## References

1. Zack, T. I. et al. Pan-cancer patterns of somatic copy number alteration. *Nat. Genet.* **45**, 1134–1140 (2013).
2. Bielski, C. M. et al. Genome doubling shapes the evolution and prognosis of advanced cancers. *Nat. Genet.* **50**, 1189–1195 (2018).
3. López, S. et al. Interplay between whole-genome doubling and the accumulation of deleterious alterations in cancer evolution. *Nat. Genet.* **52**, 283–293 (2020).
4. Storchova, Z. & Kuffer, C. The consequences of tetraploidy and aneuploidy. *J. Cell Sci.* **121**, 3859–3866 (2008).
5. Storchova, Z. & Pellman, D. From polyploidy to aneuploidy, genome instability and cancer. *Nat. Rev. Mol. Cell Biol.* **5**, 45–54 (2004).
6. Dewhurst, S. M. et al. Tolerance of whole- genome doubling propagates chromosomal instability and accelerates cancer genome evolution. *Cancer Discov.* **4**, 175–185 (2014).
7. Fox, D. T., Gall, J. G. & Spradling, A. C. Error-prone polyploid mitosis during normal *Drosophila* development. *Genes Dev.* **24**, 2294–2302 (2010).
8. Goupil, A. et al. Chromosomes function as a barrier to mitotic spindle bipolarity in polyploid cells. *J. Cell Biol.* **219**, e201908006 (2020).
9. Fujiwara, T. et al. Cytokinesis failure generating tetraploids promotes tumorigenesis in p53-null cells. *Nature* **437**, 1043–1047 (2005).
10. Orr-Weaver, T. L. When bigger is better: the role of polyploidy in organogenesis. *Trends Genet.* **31**, 307–315 (2015).
11. Gemble, S. & Basto, R. CHRONOCRISIS: when cell cycle asynchrony generates DNA damage in polyploid cells. *BioEssays* **42**, e2000105 (2020).
12. Zeman, M. K. & Cimprich, K. A. Causes and consequences of replication stress. *Nat. Cell Biol.* **16**, 2–9 (2014).

13. Koundrioukoff, S. et al. Stepwise activation of the ATR signaling pathway upon increasing replication stress impacts fragile site integrity. *PLoS Genet.* **9**, e1003643 (2013).
14. Panier, S. & Boulton, S. J. Double-strand break repair: 53BP1 comes into focus. *Nat. Rev. Mol. Cell Biol.* **15**, 7–18 (2014).
15. Ganem, N. J. et al. Cytokinesis failure triggers hippo tumor suppressor pathway activation. *Cell* **158**, 833–848 (2014).
16. Zhao, B., Rothenberg, E., Ramsden, D. A. & Lieber, M. R. The molecular basis and disease relevance of non-homologous DNA end joining. *Nat. Rev. Mol. Cell Biol.* **21**, 765–781 (2020).
17. Bester, A. C. et al. Nucleotide deficiency promotes genomic instability in early stages of cancer development. *Cell* **145**, 435–446 (2011).
18. Burgess, A., Lorca, T. & Castro, A. Quantitative live imaging of endogenous DNA replication in mammalian cells. *PLoS ONE* **7**, e45726 (2012).
19. Michalet, X. et al. Dynamic molecular combing: stretching the whole human genome for high- resolution studies. *Science* **277**, 1518–1523 (1997).
20. Zhong, Y. et al. The level of origin firing inversely affects the rate of replication fork progression. *J. Cell Biol.* **201**, 373–383 (2013).
21. Tatsumi, Y., Ohta, S., Kimura, H., Tsurimoto, T. & Obuse, C. The ORC1 cycle in human cells: I. Cell cycle-regulated oscillation of human ORC1. *J. Biol. Chem.* **278**, 41528–41534 (2003).
22. Remus, D. et al. Concerted loading of Mcm2–7 double hexamers around DNA during DNA replication origin licensing. *Cell* **139**, 719–730 (2009).
23. Siddiqui, K., On, K. F. & Diffley, J. F. X. Regulating DNA replication in Eukarya. *Cold Spring Harb. Perspect. Biol.* **5**, a012930 (2013).
24. Yan, Z. et al. Cdc6 is regulated by E2F and is essential for DNA replication in mammalian cells. *Proc. Natl Acad. Sci. USA* **95**, 3603–3608 (1998).
25. Moyer, S. E., Lewis, P. W. & Botchan, M. R. Isolation of the Cdc45/Mcm2–7/GINS (CMG) complex, a candidate for the eukaryotic DNA replication fork helicase. *Proc. Natl Acad. Sci. USA* **103**, 10236–10241 (2006).

26. Kumagai, A., Shevchenko, A., Shevchenko, A. & Dunphy, W. G. Treslin collaborates with TopBP1 in triggering the initiation of DNA replication. *Cell* **140**, 349–359 (2010).
27. DeGregori, J., Kowalik, T. & Nevins, J. R. Cellular targets for activation by the E2F1 transcription factor include DNA synthesis- and G1/S-regulatory genes. *Mol. Cell Biol.* **15**, 4215–4224 (1995).
28. Pardee, A. B. G1 events and regulation of cell proliferation. *Science* **246**, 603–608 (1989).
29. Cross, F. R., Buchler, N. E. & Skotheim, J. M. Evolution of networks and sequences in eukaryotic cell cycle control. *Phil. Trans. R. Soc. B* **366**, 3532–3544 (2011).
30. Bertoli, C., Skotheim, J. M. & De Bruin, R. A. M. Control of cell cycle transcription during G1 and S phases. *Nat. Rev. Mol. Cell Biol.* **14**, 518–528 (2013).
31. Cadart, C. et al. Size control in mammalian cells involves modulation of both growth rate and cell cycle duration. *Nat. Commun.* **9**, 3275 (2018).
32. Nano, M. et al. Cell-cycle asynchrony generates DNA damage at mitotic entry in polyploid cells. *Curr. Biol.* **29**, 3937–3945.e7 (2019).
33. Frawley, L. E. & Orr-Weaver, T. L. Polyploidy. *Curr. Biol.* **25**, R353–R358 (2015).
34. Carter, S. L. et al. Absolute quantification of somatic DNA alterations in human cancer. *Nat. Biotechnol.* **30**, 413–421 (2012).
35. Taylor, A. M. et al. Genomic and functional approaches to understanding cancer aneuploidy. *Cancer Cell* **33**, 676–689.e3 (2018).
36. Hégarat, N. et al. Cyclin A triggers mitosis either via the greatwall kinase pathway or cyclin B. *EMBO J.* **39**, e104419 (2020).
37. SB, K. et al. A quantitative FastFUCCI assay defines cell cycle dynamics at a single-cell level. *J. Cell Sci.* **130**, 512–520 (2017).
38. KS, Y., RH, K., M, L., R, G. & R, W. Single cell resolution *in vivo* imaging of DNA damage following PARP inhibition. *Sci Rep.* **5**, 10129 (2015).

39. Karess, R. E. et al. The regulatory light chain of nonmuscle myosin is encoded by *spaghetti-squash*, a gene required for cytokinesis in *Drosophila*. *Cell* **65**, 1177–1189 (1991).
40. Gemble, S. et al. Centromere dysfunction compromises mitotic spindle pole integrity. *Curr. Biol.* **29**, 3072–3080.e5 (2019).
41. J, L., BO, P., K, H., J, B. & K, H. Deregulated expression of E2F family members induces S-phase entry and overcomes p16INK4A-mediated growth suppression. *Mol. Cell. Biol.* **16**, 1047–1057 (1996).
42. Sakaue-Sawano, A. et al. Visualizing spatiotemporal dynamics of multicellular cell-cycle progression. *Cell* **132**, 487–498 (2008).
43. Aknoun, S. et al. Living cell dry mass measurement using quantitative phase imaging with quadriwave lateral shearing interferometry: an accuracy and sensitivity discussion. *J. Biomed. Opt.* **20**, 126009 (2015).
44. Bon, P., Maucort, G., Wattellier, B. & Monneret, S. Quadriwave lateral shearing interferometry for quantitative phase microscopy of living cells. *Opt. Express* **17**, 13080 (2009).
45. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
46. Jun, G., Wing, M. K., Abecasis, G. R. & Kang, H. M. An efficient and scalable analysis framework for variant extraction and refinement from population-scale DNA sequence data. *Genome Res.* **25**, 918–925 (2015).
47. Bakker, B. et al. Single-cell sequencing reveals karyotype heterogeneity in murine and human malignancies. *Genome Biol.* **17**, 115 (2016).
48. van den Bos, H. et al. Single-cell whole genome sequencing reveals no evidence for common aneuploidy in normal and Alzheimer's disease neurons. *Genome Biol.* **17**, 116 (2016).
49. Subramanian, A. et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl Acad. Sci. USA* **102**, 15545–15550 (2005).
50. Mootha, V. K. et al. PGC-1 $\alpha$ -responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nat. Genet.* **34**, 267–273 (2003).

## Acknowledgements

We authors acknowledge the Cell and Tissue Imaging platform (PICT-IBiSA), member of the French National Research Infrastructure France-BioImaging (ANR10-INBS-04) and the Nikon Imaging center from Institut Curie for microscopy. We thank L. Guyonnet, A. Chipont, C. Guerrin and C. Roulin from the Cytometry and RADEXP platforms of Institut Curie; and V. Marthiens, S. Lambert, E. Schwob, W. Marshall, G. Almouzni, J. S. Hoffmann, D. Fachinetti, M. Budzyk, F. Edwards, G. Fantozzi, R. Salamé, A. Goupil, C. Chen, F. Perez and N. Ganem for helpful discussions and/or comments on the manuscript. This work was supported by FOR2800/STO918-7 to Z.S., ERC CoG (ChromoNumber-725907) to R.B., Institut Curie and the CNRS. The Basto laboratory is a member of the Cell(n)Scale Labex.

## Author information

### Author notes

1. Maddalena Nano

Present address: Molecular, Cellular, and Developmental Biology Department, University of California, Santa Barbara, CA, USA

2. These authors contributed equally: René Wardenaar, Kristina Keuper

## Affiliations

1. Institut Curie, PSL Research University, CNRS, UMR144, Biology of Centrosomes and Genetic Instability Laboratory, Paris, France

Simon Gemble, Maddalena Nano, Anthony Simon, Oumou Goundiam & Renata Basto

2. European Research Institute for the Biology of Ageing, University of Groningen, University Medical Center Groningen, Groningen, The Netherlands

René Wardenaar, Andréa E. Tijhuis, Diana C. J. Spierings & Floris Foijer

3. Department of Molecular Genetics, TU Kaiserslautern, Kaiserslautern, Germany

Kristina Keuper, Sara Vanessa Bernhard & Zuzana Storchová

4. Institut Curie and Institut Pierre Gilles de Gennes, PSL Research University, CNRS, UMR 144, Systems Biology of Cell Polarity and Cell Division, Paris, France

Nishit Srivastava & Matthieu Piel

5. Cell and Tissue Imaging Facility (PICT-IBiSA), Institut Curie, PSL Research University, Centre National de la Recherche Scientifique, Paris, France

Anne-Sophie Macé

6. Genome Damage and Stability Centre, School of Life Sciences, University of Sussex, Brighton, UK

Helfrid Hochegger

## Contributions

S.G. and R.B. conceived the project and wrote the manuscript. DNA damage characterization and the analysis of its origins were conceived together with Z.S. S.G. did most of the experiments and data analysis presented here. M.N. did the initial observations of high levels of DNA damage in *Drosophila* polyploid neuroblasts. R.W., A.E.T., D.C.J.S. and F.F. did the scDNA-seq and bio-informatical analysis. K.K., S.V.B. and Z.S. contributed to DNA combing. A.-S.M. helped with image quantifications and analysis. A.S. performed the 3D spheroids experiments and helped with image quantification. O.G. performed the bio-informatical analysis of ovarian TCGA tumours. N.S. and M.P. performed the quantitative phase imaging experiments and analysis and H.H. contributed with unpublished cell lines. All authors read and comment on the manuscript.

## Corresponding authors

Correspondence to [Simon Gemble](#) or [Renata Basto](#).

## Ethics declarations

### Competing interests

The authors declare no competing interests.

## Peer review

## Peer review information

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

## Additional information

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## Extended data figures and tables

### Extended Data Fig. 1 Characterization of RPE-1 cells upon WGD.

**(a, d and g)** Graphs showing the percentage of tetraploid interphase RPE-1 cells in the indicated experimental conditions. Mean  $\pm$  SEM, > 100 interphase cells were analyzed from three independent experiments. **(b, e and h)** Graphs showing the percentage of mono- and multinucleated RPE-1 tetraploid cells in the indicated experimental conditions. Mean  $\pm$  SEM, > 100 interphase cells were analyzed from three independent experiments. **(c, f and i)** Graphs representing the nuclear area in diploid (D) and tetraploid (T) RPE-1 cells. Mean  $\pm$  SEM, >100 interphase cells were analyzed from three independent experiments. **(j)** Graph showing the correlation between the number of  $\gamma$ H2AX foci and  $\gamma$ H2AX foci intensity in diploid (left panel, gray) and tetraploid (right panel, blue) RPE-1 cells induced through MS. >100 interphase cells were analyzed from three independent experiments. **(k–l)** Graphs showing the number of  $\gamma$ H2AX foci relative to nuclear area (**k**) or DAPI fluorescence intensity (**l**) in diploid (gray) and tetraploid (blue) RPE-1 cells induced through MS. t-test (two-sided) **(a, d and g)**. ANOVA test (one-sided) **(c, f, i, k and l)**. Pearson test (two-sided) **(j)**.

### Source data

### Extended Data Fig. 2 Additional methods and cell lines confirm that WGD generates high levels of DNA damage within the first interphase.

**(a, c and e)** Images showing diploid and tetraploid (generated as indicated) RPE-1 cells labeled with  $\gamma$ H2AX (red) and  $\beta$ -Catenin (gray) antibodies. DNA in blue. **(b, d and f)** Graphs showing the number of  $\gamma$ H2AX foci in diploid (D) and tetraploid (T) RPE-1. Mean  $\pm$  SEM, >100 interphase cells were analyzed from at least three independent experiments. **(g–h)** Left - Graph showing the percentage of tetraploid interphase cells in BJ **(g)** or HCT116 **(h)** cell lines. Mean  $\pm$  SEM, >100 interphase cells were analyzed from at least three independent experiments. Right - Graph representing

the number of  $\gamma$ H2AX foci in diploid and tetraploid BJ (**g**) or HCT116 (**h**) cells. Mean  $\pm$  SEM, >100 interphase cells were analyzed from at least three independent experiments. (**i**) Graph showing the number of  $\gamma$ H2AX foci in diploid (gray) or tetraploid (blue) RPE-1 cells treated with 1  $\mu$ M APH or 2 mM HU. Mean  $\pm$  SEM, >100 interphase cells were analyzed from at least three independent experiments. (**j**) Comet images from diploid (left) and tetraploid (right) RPE-1 cells. (**k**) Graph showing the olive moment in diploid and tetraploid RPE-1 (left) or BJ (right) cell lines. Mean  $\pm$  SEM, > 100 comets were analyzed from two independent experiments. (**l**) p53 and tubulin levels assessed by western blot. Etoposide was added as a control for the increased p53 levels. (**m** and **o**) Graphs representing the mean number of  $\gamma$ H2AX foci per interphase cell over time (days in culture) in diploid and tetraploid RPE-1 cells. Mean  $\pm$  SEM, > 100 interphase cells were analyzed from two independent experiments. The dotted lines indicate nuclear area. ANOVA test (one-sided) (**b**, **d**, **f**, **g**, **h**, **i**, **k**, **m** and **o**).

[Source data](#)

[\*\*Extended Data Fig. 3 DNA damage is generated during the first S-phase upon WGD.\*\*](#)

(**a**) 3D RPE-1 spheroid low magnification (top) and insets of two cells showing diploid and tetraploid nuclei (bottom) induced through MS labeled with  $\gamma$ H2AX (red) and  $\beta$ -Catenin (yellow) antibodies. DNA in blue. (**b–d**)  $\gamma$ H2AX index in diploid and tetraploid RPE-1 (**b**), BJ (**c**) and HCT116 (**d**) spheroids. Mean  $\pm$  SEM, > 95 interphase cells were analyzed from at least two independent experiments. (**e–f**)  $\gamma$ H2AX foci in diploid and tetraploid RPE-1 cells over time. (**g**) Left - Stills of RPE-1 cells expressing RFP-H2B and GFP-53BP1 time lapse videos. Right- 53BP1 foci number in diploid and tetraploid cells. Mean  $\pm$  SEM, > 40 interphase cells were analyzed from three independent experiments. (**h**) 53BP1 foci number in fixed diploid and tetraploid RPE-1. (**i**) Cell cycle distribution of RPE-1 cells in the indicated conditions. (**j**) Percentage of RPE-1 cells in G1, S and G2-M in the indicated conditions. Mean  $\pm$  SEM, >30 000 cells from at least three independent experiments. (**k**) Workflow used to analyze G1 or S-phase cells. (**l** and **m**)  $\gamma$ H2AX foci number in diploid and tetraploid RPE-1 cells as indicated. Experiments (**l** and **m**) share the same reference control. (**n**, **o**)  $\gamma$ H2AX foci number in diploid and tetraploid RPE-1 cells synchronized in G1 using 0,5  $\mu$ M abemaciclib (**n**) or 1  $\mu$ M K03861 (**o**) or released in S-phase. (**p**) Images of  $\gamma$ H2AX (red) and EdU (cyan)/ PCNA (yellow) foci co-localization in S-phase in diploid and tetraploid RPE-1 cells. DNA in blue. White squares highlight higher magnifications. (**q**) Percentage of replication sites (EdU) colocalizing with  $\gamma$ H2AX foci. Mean  $\pm$  SEM, >50 interphase cells were analyzed from at least three independent experiments. For (**e**, **f**, **h**, **l–o**) Mean  $\pm$  SEM, >100 interphase cells were analyzed from at least three

independent experiments. The dotted lines indicate nuclear area. ANOVA test (one-sided) (**b**, **c**, **d**, **e**, **f**, **g**, **h**, **j**, **l**, **m**, **n** and **o**). t-test (two-sided) (**q**).

[Source data](#)

**Extended Data Fig. 4 DNA damage in newly born tetraploid S-phase cells is associated with HR and RS-associated markers.**

**(a and c)** Diploid and tetraploid RPE-1 S-phase cells labeled with XRCC1 (**a**, in green) or KU80 antibodies (**c**, in green). DNA in blue. **(b and d)** Number of XRCC1 (**b**) or KU80 (**d**) foci in diploid and tetraploid RPE-1 cells synchronized in G1 using 1 $\mu$ M palbociclib or released in S-phase. Mean  $\pm$  SEM, >100 interphase cells were analyzed from at least three independent experiments. **(e)** Images showing  $\gamma$ H2AX (red) and RAD51 (yellow) foci colocalization (white arrows) in diploid and tetraploid RPE-1 cells. DNA in blue. The white squares correspond to higher magnification regions. **(f)** RAD51 foci number in diploid and tetraploid RPE-1 cells arrested in G1 using 1  $\mu$ M palbociclib or released in S-phase. Mean  $\pm$  SEM, >100 interphase cells analyzed from at least three independent experiments. **(g)** Percentage of colocalizing  $\gamma$ H2AX and RAD51 signals in diploid and tetraploid RPE-1 cells arrested in G1 using 1  $\mu$ M palbociclib or released in S-phase. Mean  $\pm$  SEM, >50 interphase cells were analyzed from at least three independent experiments. **(h)** Images showing the colocalization (white arrows) of  $\gamma$ H2AX (red) and FANCD2 (yellow). DNA in blue. **(i)** FANCD2 foci number in diploid and tetraploid RPE-1 cells. Mean  $\pm$  SEM, >80 interphase cells were analyzed from at least three independent experiments. **(j)** Graph representing the percentage of  $\gamma$ H2AX signal colocalizing with FANCD2 foci in diploid and tetraploid RPE-1 interphase cells. Mean  $\pm$  SEM, >50 interphase cells were analyzed from at least three independent experiments. **(k)** Graph showing RPA number foci in diploid and tetraploid RPE-1 cells arrested in G1 using 1  $\mu$ M palbociclib or released in S-phase. Mean  $\pm$  SEM, >100 interphase cells were analyzed from at least three independent experiments. The dotted lines indicate the nuclear area. ANOVA test (two sided) (**b**, **d**, **f**, **g**, **i**, **j** and **k**).

[Source data](#)

**Extended Data Fig. 5 DNA damage in newly born tetraploid cells is generated in a DNA replication-dependent manner.**

**(a)**  $\gamma$ H2AX foci number in diploid and tetraploid RPE-1 cells released in S-phase  $\pm$  1  $\mu$ M PHA. Mean  $\pm$  SEM, >100 interphase cells were analyzed from at least three independent experiments. **(b–c)**  $\gamma$ H2AX foci number in diploid and tetraploid RPE-1 cells, arrested in G1 using 1  $\mu$ M palbociclib or released in S-phase  $\pm$  400 nM aphidicolin (APH) (**b**) or 1  $\mu$ M PHA (**c**). Mean  $\pm$  SEM, >100 interphase cells were

analyzed from at least three independent experiments. **(d)**  $\gamma$ H2AX foci number in diploid and tetraploid RPE-1 cells released in S-phase  $\pm$  400 nM APH. Mean  $\pm$  SEM, >100 interphase cells were analyzed from at least three independent experiments. **(e–f)**  $\gamma$ H2AX foci number in diploid and tetraploid BJ **(e)** or HCT116 **(f)** cells, released in S-phase  $\pm$  400 nM APH. Mean  $\pm$  SEM, >100 interphase cells were analyzed from at least three independent experiments. **(g)** Images showing EdU  $\pm$  tetraploid RPE-1 cells.  $\gamma$ H2AX antibodies in red, EdU in yellow and DNA in blue. **(h)**  $\gamma$ H2AX foci number relative to EdU intensity in RPE-1 tetraploid cells released in S-phase untreated (left panel) or treated (right panel) with 400 nM aphidicolin (APH). Mean  $\pm$  SEM, >100 interphase cells were analyzed from at least three independent experiments. **(i, j)**  $\gamma$ H2AX foci number in diploid and tetraploid cells **(i, blue)** or EnR **(j, red)**, synchronized in G1 using 1  $\mu$ M palbociclib or released in S-phase  $\pm$  nucleosides at two different concentrations (methods). Mean  $\pm$  SEM, >100 interphase cells were analyzed from at least three independent experiments. The dotted lines indicate the nuclear area. ANOVA test (one-sided) **(a, b, c, d, e, f, i and j)**. Pearson test (two-sided) **(h)**.

#### [Source data](#)

#### [\*\*Extended Data Fig. 6 DNA replication dynamics is impaired during the first S-phase in tetraploid cells.\*\*](#)

**(a)** Percentage of cells per cell cycle phase in RPE-1 (dark gray) and RPE-1 PCNA<sup>chromo</sup> cell lines (light gray). **(b)** Workflow depicting methods used to process and analyze DNA replication by time-lapse. **(c)** Stills of time lapse movies of diploid and tetraploid RPE-1 PCNA<sup>chromo</sup> cells. Active replication sites are visualized using PCNA chromobodies (in cyan) and reconstructed using *Imaris* in 3D (in red). **(d)** Total number of active replication sites in S-phase in diploid and tetraploid RPE-1 cells. Mean  $\pm$  SEM, >20 S-phase cells were analyzed from three independent experiments. **(e)** EdU foci number relative to nuclear area in diploid and tetraploid RPE-1 cells in mid (T5) or late (T9) S-phase. Mean  $\pm$  SEM, >100 interphase cells were analyzed from at least three independent experiments. **(f)** Volume of active replication sites (in  $\mu\text{m}^3$ ) for diploid and tetraploid RPE-1 PCNA<sup>chromo</sup> cells. Mean  $\pm$  SEM, at least 1000 active replication sites were analyzed from three independent experiments. **(g)** Mean number of active replication sites over time in diploid and tetraploid RPE-1 cells. >20 S-phase cells were analyzed from two independent experiments (see Supplementary Data 1). **(h)** Ratio of early/late S-phase duration in diploid or tetraploid RPE-1 PCNA<sup>chromo</sup> cells  $\pm$  extended G1 duration. Mean  $\pm$  SEM, > 70 cells from two independent experiments were analyzed. **(i)** S-phase duration in diploid or tetraploid RPE-1 PCNA<sup>chromo</sup> cells  $\pm$  extended G1 duration. Mean  $\pm$  SEM, > 70 cells from two independent experiments were analyzed. **(j)** Replication fork speed in diploid and tetraploid HCT116 cells. Mean  $\pm$  SEM, > 250 replication forks were analyzed. **(k)**

Proportion of fibers with the indicated inter-origin distance (kb) in diploid or tetraploid HCT116 cells. Mean  $\pm$  SEM,  $> 75$  replication origins were analyzed. ANOVA test (one-sided) (**a** and **e**). t-test (two-sided) (**d**, **f**, **h**, **i**, **j** and **k**).

[Source data](#)

**Extended Data Fig. 7 Genome wide analysis of RPE-1 and BJ tetraploid cells.**

(**a, b**) Genome-wide copy number plots of G1 and G2/M diploid RPE-1 cells and G1 tetraploid RPE-1 cells (**a**) were generated using the standard version of the Aneufinder algorithm and genome-wide copy number plots of G1 tetraploid BJ cells (**b**) were generated using a modified version of the Aneufinder algorithm (see [methods](#)). G2/M conditions were normalized using G1 cells. Each row represents a cell. The copy number state (in 1-Mb bins) is indicated in color (with aberrations contrasting from green in diploid G1 (2n) or from yellow in diploid G2/M or tetraploid G1 (4n)). (**c**) Table showing aneuploidy and heterogeneity scores in the indicated conditions. (**d**) Graph showing the number of aneuploid chromosomes per cell in the diploid G1 and G2/M (in gray) and in tetraploid G1 and G2/M (in blue) cells. The percentage of cells with  $\geq 1$  aneuploid chromosome is indicated under the graph. ANOVA test (one-sided) (**d**, left panel). t-test (two-sided) (**d**, right panel).

[Source data](#)

**Extended Data Fig. 8 DNA damage analysis in 3D cultures.**

(**a**) Left panel - Representative cell cycle distribution of diploid and tetraploid RPE-1 cells. Right panels - RNA content in diploid (in gray) and tetraploid (in blue) populations. (**b, c**) Graphs showing the relative RNA levels in diploid (D, in gray) and tetraploid (T) cells generated through MS (**b**, blue) or EnR (**c**, red). (**d**) Representative images of cell sorting experiments according to cell cycle stage (RFP+ for G1 cells) and DNA content. (**e**) Graph showing the percentage of interphase tetraploid cells in diploid (gray) and tetraploid (blue) RPE-1 cell populations obtained after cell sorting. Mean  $\pm$  SEM,  $> 100$  interphase cells from at least three independent experiments were analyzed. (**f, g**) Graphs representing actin (**f**) and  $\beta$ -Catenin (**g**) levels relative to H2B levels (fold change) in total protein extracts from diploid (gray) and tetraploid (blue) cells. Mean  $\pm$  SEM from at least three independent experiments. (**h**) Graph showing H2B levels in the chromatin bound fraction in diploid (gray) and tetraploid (blue) cells. Mean  $\pm$  SEM from at least three independent experiments. (**i, j**) Graph representing G1 duration in diploid (gray) or tetraploid cells generated through MS (**i**, blue) or EnR (**j**, red). The dotted lines indicate the nuclear area. The white squares correspond to higher magnification. t-test (one-sided) (**b, c, e, f, g, h, i** and **j**).

[Source data](#)

**[Extended Data Fig. 9 G1 lengthening restores DNA replication dynamics and results in a decrease in the levels of DNA damage in tetraploid cells.](#)**

**(a, b)** RPE-1 cell cycle profile and percentage of cells in the indicated conditions. Mean  $\pm$  SEM,  $> 30\,000$  cells from at least three independent experiments. **(c)** RPE FUCCI in diploid and tetraploid cells treated with 160 nM palbociclib. **(d)** Stills of time lapse videos of diploid and tetraploid RPE-1 PCNA<sup>chromo</sup> cells with extended G1. Active replication sites visualized using PCNA chromobodies (cyan) and reconstructed using *Imaris* in 3D (red). **(e)** Active replication sites average number over time with extended G1. Mean  $\pm$  SEM,  $> 11$  S-phase cells analyzed, two independent experiments (see Supplementary Data 1). **(f)** Active replication sites total number with extended G1. Mean  $\pm$  SEM,  $> 11$  S-phase cells were analyzed, two independent experiments. **(g)** Active replication sites volume ( $\mu\text{m}^3$ ) with extended G1. Mean  $\pm$  SEM,  $> 1000$  Active replication sites analyzed, three independent experiments. **(h)** EdU foci number relative to nuclear area with extended G1. Mean  $\pm$  SEM,  $> 100$  interphase cells, at least three independent experiments. **(i)** Ratio of early/late S phase duration  $\pm$  extended G1. Mean  $\pm$  SEM,  $> 70$  cells, two independent experiments. **(j)** S-phase duration  $\pm$  extended G1. Mean  $\pm$  SEM,  $> 70$  cells, two independent experiments. **(k)** H2B levels in chromatin bound extracts. Mean  $\pm$  SEM, four independent experiments. **(l and m)** FANCD2 or 53BP1 foci number in cells synchronized in G1 or released in S-phase  $\pm$  extended G1. **(n–p)**  $\gamma$ H2AX foci number in cells synchronized in G1 using the indicated treatments or released in S-phase  $\pm$  extended G1. **(q and r)**  $\gamma$ H2AX foci number in diploid and tetraploid BJ (q) or HCT116 (r) cells synchronized in G1 or released in S-phase  $\pm$  extended G1. **(l–r)** Mean  $\pm$  SEM,  $> 100$  interphase cells were analyzed, at least three independent experiments. The dotted lines indicate the nuclear area. ANOVA test (one-sided) (b, h, i, j, l, m, n, o, p, q and r). T-test (two-sided) (f, g and k).

[Source data](#)

**[Extended Data Fig. 10 E2F1OE decreases DNA damage levels in tetraploid human cell lines and in \*Drosophila\* NBs.](#)**

**(a)** Western blot documenting the levels of E2F1 and tubulin from cell lysates obtained from diploid RPE-1 cells  $\pm$  E2F1-HA over-expression (OE). **(b–c)** Graphs representing the number of  $\gamma$ H2AX foci per interphase cell in diploid (D) and tetraploid (T) BJ (b) or HCT116 (c) cells released in S-phase  $\pm$  E2F1 OE. Mean  $\pm$  SEM,  $> 100$  interphase cells were analyzed from at least three independent experiments. **(d)** Graph showing wild type salivary gland cell (in gray), diploid (in gray) and polyplloid NBs (in yellow) area ( $\text{in } \mu\text{m}^2$ ). Mean  $\pm$  SEM,  $> 60$  cells were

analyzed per condition. **(e)** Graph showing  $\gamma$ H2Av indexes in diploid (in gray) or polyploid NBs (in yellow) induced through CF by depleting Pavarotti. Mean  $\pm$  SEM, >40 cells were analyzed per condition. **(f)** Graph showing the cell area ( $\mu\text{m}^2$ ) of diploid (gray) and polyploid NBs (yellow)  $\pm$  E2F1OE. Mean  $\pm$  SEM, >30 cells were analyzed per condition. **(g)** Graph representing the  $\gamma$ H2Av index in polyploid NBs  $\pm$  10 $\mu\text{M}$  nucleosides. Mean  $\pm$  SEM, >28 cells were analyzed per condition. **(h)** Gene set enrichment analysis from GSEA. Plots show significant enrichment of DNA repair genes in near-tetraploid tumors when compared to near-diploid tumors in lung, bladder and ovarian cancers (TCGA pan cancer data set). **(h)** p value from false discovery rate (FDR; methods). ANOVA test (one-sided) **(b, c, d, f)**. t-test (two-sided) **(e and g)**.

## Source data

# Supplementary information

## Supplementary Information

This file contains Supplementary Discussion, Methods, Data 1, 2, Fig. 1, references and legends for Supplementary Videos

## Reporting Summary

## Peer Review File

## Supplementary Video 1

DNA damage in tetraploid is generated during S phase

## Supplementary Video 2

DNA damage in tetraploid is generated during S phase

## Supplementary Video 3

Quantitative 4D live imaging of endogenous DNA replication

## Supplementary Video 4

Quantitative 4D live imaging of endogenous DNA replication

## Supplementary Video 5

Quantitative 4D live imaging of endogenous DNA replication

## **Supplementary Video 6**

Quantitative 4D live imaging of endogenous DNA replication

## **Source data**

**Source Data Fig. 1**

**Source Data Fig. 2**

**Source Data Fig. 3**

**Source Data Fig. 4**

**Source Data Extended Data Fig. 1**

**Source Data Extended Data Fig. 2**

**Source Data Extended Data Fig. 3**

**Source Data Extended Data Fig. 4**

**Source Data Extended Data Fig. 5**

**Source Data Extended Data Fig. 6**

**Source Data Extended Data Fig. 7**

**Source Data Extended Data Fig. 8**

**Source Data Extended Data Fig. 9**

**Source Data Extended Data Fig. 10**

## **Rights and permissions**

**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

Gamble, S., Wardenaar, R., Keuper, K. *et al.* Genetic instability from a single S phase after whole-genome duplication. *Nature* **604**, 146–151 (2022).  
<https://doi.org/10.1038/s41586-022-04578-4>

- Received: 16 July 2021
- Accepted: 23 February 2022
- Published: 30 March 2022
- Issue Date: 07 April 2022
- DOI: <https://doi.org/10.1038/s41586-022-04578-4>

### Share this article

Anyone you share the following link with will be able to read this content:

[Get shareable link](#)

Sorry, a shareable link is not currently available for this article.

[Copy to clipboard](#)

Provided by the Springer Nature SharedIt content-sharing initiative

## Further reading

- [Genome doubling causes double trouble](#)

- Yonatan Eliezer
- Uri Ben-David

*Nature* (2022)

### [Genome doubling causes double trouble](#)

- Yonatan Eliezer
- Uri Ben-David

News & Views 30 Mar 2022

---

This article was downloaded by **calibre** from <https://www.nature.com/articles/s41586-022-04578-4>

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

- Article
- [Published: 30 March 2022](#)

# Crucial role and mechanism of transcription-coupled DNA repair in bacteria

- [Binod K. Bharati<sup>1,2</sup> na1](#),
- [Manjunath Gowder<sup>1,2</sup> na1](#),
- [Fangfang Zheng<sup>3</sup>](#),
- [Khaled Alzoubi<sup>1</sup>](#),
- [Vladimir Svetlov<sup>1</sup>](#),
- [Venu Kamarthapu<sup>1,2</sup>](#),
- [Jacob W. Weaver<sup>1</sup>](#),
- [Vitaly Epshtain<sup>1</sup>](#),
- [Nikita Vasilyev ORCID: orcid.org/0000-0001-9918-5735<sup>1</sup>](#),
- [Liqiang Shen<sup>3</sup>](#),
- [Yu Zhang ORCID: orcid.org/0000-0002-1778-8389<sup>3</sup>](#) &
- [Evgeny Nudler ORCID: orcid.org/0000-0002-8811-3071<sup>1,2</sup>](#)

[Nature](#) volume 604, pages 152–159 (2022)

- 3800 Accesses
- 375 Altmetric
- [Metrics details](#)

## Subjects

- [Nucleotide excision repair](#)

- [Transcription](#)

## Abstract

Transcription-coupled DNA repair (TCR) is presumed to be a minor sub-pathway of nucleotide excision repair (NER) in bacteria. Global genomic repair is thought to perform the bulk of repair independently of transcription. TCR is also believed to be mediated exclusively by Mfd—a DNA translocase of a marginal NER phenotype<sup>1,2,3</sup>. Here we combined in cellulo cross-linking mass spectrometry with structural, biochemical and genetic approaches to map the interactions within the TCR complex (TCRC) and to determine the actual sequence of events that leads to NER *in vivo*. We show that RNA polymerase (RNAP) serves as the primary sensor of DNA damage and acts as a platform for the recruitment of NER enzymes. UvrA and UvrD associate with RNAP continuously, forming a surveillance pre-TCRC. In response to DNA damage, pre-TCRC recruits a second UvrD monomer to form a helicase-competent UvrD dimer that promotes backtracking of the TCRC. The weakening of UvrD–RNAP interactions renders cells sensitive to genotoxic stress. TCRC then recruits a second UvrA molecule and UvrB to initiate the repair process. Contrary to the conventional view, we show that TCR accounts for the vast majority of chromosomal repair events; that is, TCR thoroughly dominates over global genomic repair. We also show that TCR is largely independent of Mfd. We propose that Mfd has an indirect role in this process: it participates in removing obstructive RNAPs in front of TCRCs and also in recovering TCRCs from backtracking after repair has been completed.

This is a preview of subscription content

## Access options

Subscribe to Nature+

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

\$29.99

monthly

[Subscribe](#)

Subscribe to Journal

Get full journal access for 1 year

\$199.00

only \$3.90 per issue

[Subscribe](#)

All prices are NET prices.

VAT will be added later in the checkout.

Tax calculation will be finalised during checkout.

Buy article

Get time limited or full article access on ReadCube.

\$32.00

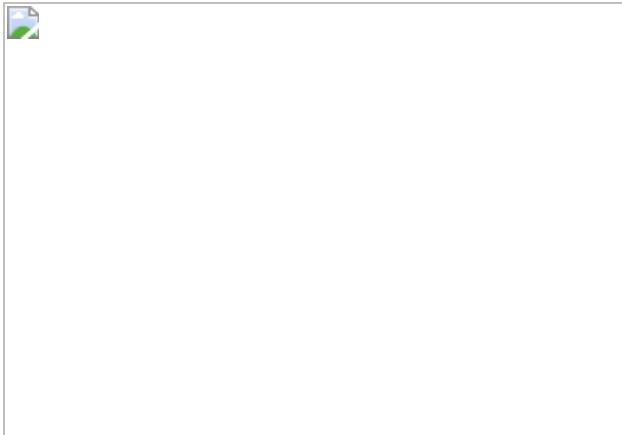
[Buy](#)

All prices are NET prices.

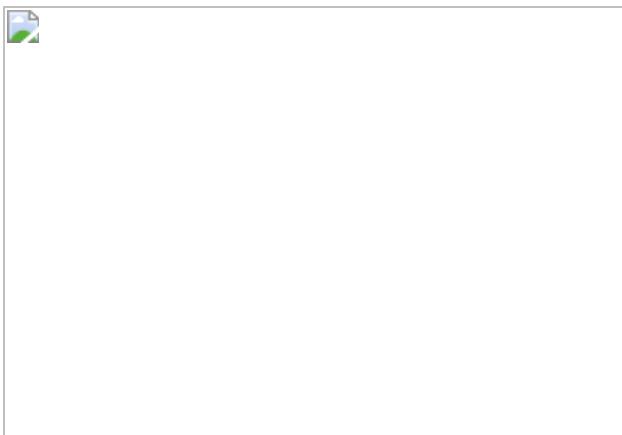
**Additional access options:**

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

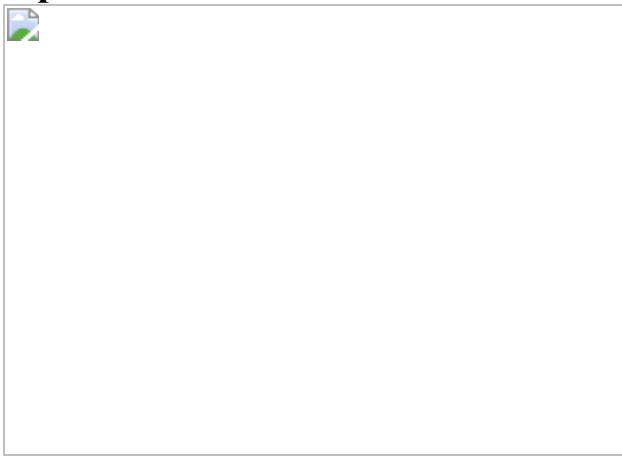
**Fig. 1: Recruitment of NER enzymes and Mfd to RNAP in vivo.**



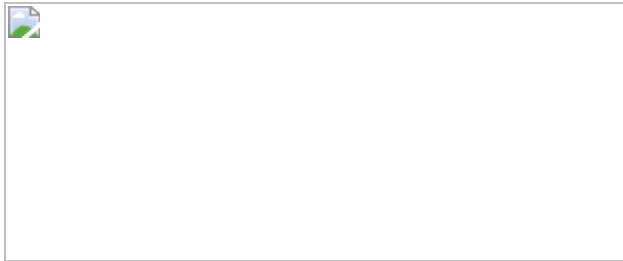
**Fig. 2: Structural organization of the pre-TCR and TCR complexes in cellulo.**



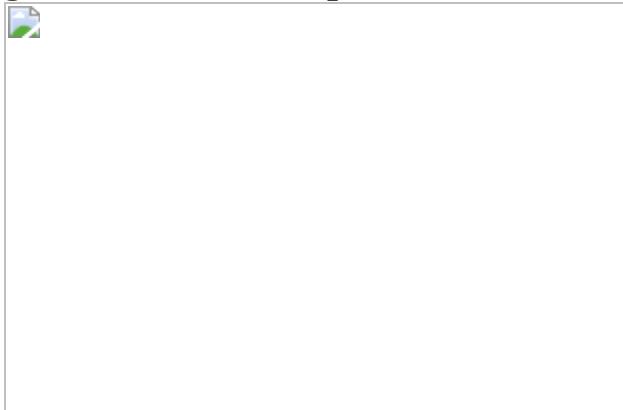
**Fig. 3: Structure and function of the UvrD CTD–RNAP  $\beta$  pincer complex.**



**Fig. 4: Global NER fully depends on ongoing transcription and UvrD, but not on Mfd.**



**Fig. 5: Local transcription enables NER (independently of Mfd).**



## Data availability

Coordinates and structure factors of the X-ray crystallography structures determined in this study have been deposited in the PDB under accession numbers [7EGT](#) and [7EGS](#), and are listed in Extended Data Table 1. PDB codes for the XLMS-driven docking models are available at [http://nudlerlab.info/media/TCRC\\_computational\\_models.zip](http://nudlerlab.info/media/TCRC_computational_models.zip).

## References

1. Sancar, A. Mechanisms of DNA repair by photolyase and excision nuclease (Nobel Lecture). *Angew. Chem. Int. Ed. Engl.* **55**, 8502–8527 (2016).
2. Adebali, O., Chiou, Y. Y., Hu, J. C., Sancar, A. & Selby, C. P. Genome-wide transcription-coupled repair in *Escherichia coli* is mediated by the Mfd translocase. *Proc. Natl Acad. Sci. USA* **114**, E2116–E2125 (2017).

3. Adebali, O., Sancar, A. & Selby, C. P. Mfd translocase is necessary and sufficient for transcription-coupled repair in *Escherichia coli*. *J. Biol. Chem.* **292**, 18386–18391 (2017).
4. Marteijn, J. A., Lans, H., Vermeulen, W. & Hoeijmakers, J. H. Understanding nucleotide excision repair and its roles in cancer and ageing. *Nat. Rev. Mol. Cell Biol.* **15**, 465–481 (2014).
5. Kisker, C., Kuper, J. & Van Houten, B. Prokaryotic nucleotide excision repair. *Cold Spring Harb. Perspect. Biol.* **5**, a012591 (2013).
6. Kuper, J. & Kisker, C. Damage recognition in nucleotide excision DNA repair. *Curr. Opin. Struct. Biol.* **22**, 88–93 (2012).
7. Hanawalt, P. C. & Spivak, G. Transcription-coupled DNA repair: two decades of progress and surprises. *Nat. Rev. Mol. Cell Biol.* **9**, 958–970 (2008).
8. Pani, B. & Nudler, E. Mechanistic insights into transcription coupled DNA repair. *DNA Repair* **56**, 42–50 (2017).
9. Spivak, G. Transcription-coupled repair: an update. *Arch. Toxicol.* **90**, 2583–2594 (2016).
10. Selby, C. P. & Sancar, A. Molecular mechanism of transcription-repair coupling. *Science* **260**, 53–58 (1993).
11. Cohen, S. E. et al. Roles for the transcription elongation factor NusA in both DNA repair and damage tolerance pathways in *Escherichia coli*. *Proc. Natl Acad. Sci. USA* **107**, 15517–15522 (2010).
12. Kamarthapu, V. et al. ppGpp couples transcription to DNA repair in *E. coli*. *Science* **352**, 993–996 (2016).
13. Ragheb, M. N. et al. Inhibiting the evolution of antibiotic resistance. *Mol. Cell* **73**, 157–165 (2019).
14. Schalow, B. J., Courcelle, C. T. & Courcelle, J. Mfd is required for rapid recovery of transcription following UV-induced DNA damage

- but not oxidative DNA damage in *Escherichia coli*. *J. Bacteriol.* **194**, 2637–2645 (2012).
15. Witkin, E. M. Radiation-induced mutations and their repair. *Science* **152**, 1345–1353 (1966).
  16. Kamarthapu, V. & Nudler, E. Rethinking transcription coupled DNA repair. *Curr. Opin. Microbiol.* **24**, 15–20 (2015).
  17. Mullenders, L. DNA damage mediated transcription arrest: step back to go forward. *DNA Repair* **36**, 28–35 (2015).
  18. Epshtain, V. et al. UvrD facilitates DNA repair by pulling RNA polymerase backwards. *Nature* **505**, 372–377 (2014).
  19. Rasouly, A., Pani, B. & Nudler, E. A magic spot in genome maintenance. *Trends Genet.* **33**, 58–67 (2017).
  20. Lin, C. G., Kovalsky, O. & Grossman, L. Transcription coupled nucleotide excision repair by isolated *Escherichia coli* membrane-associated nucleoids. *Nucleic Acids Res.* **26**, 1466–1472 (1998).
  21. Manelyte, L., Kim, Y. I., Smith, A. J., Smith, R. M. & Savery, N. J. Regulation and rate enhancement during transcription-coupled DNA repair. *Mol. Cell* **40**, 714–724 (2010).
  22. Wang, C. Y. et al. Structural basis of transcription-translation coupling. *Science* **369**, 1359–1365 (2020).
  23. Jia, H. F. et al. Rotations of the 2B sub-domain of *E. coli* UvrD helicase/translocase coupled to nucleotide and DNA binding. *J. Mol. Biol.* **411**, 633–648 (2011).
  24. Lee, J. Y. & Yang, W. UvrD helicase unwinds DNA one base pair at a time by a two-part power stroke. *Cell* **127**, 1349–1360 (2006).
  25. Pakotiprapha, D. et al. Crystal structure of *Bacillus stearothermophilus* UvrA provides insight into ATP-modulated dimerization, UvrB interaction, and DNA binding. *Mol. Cell* **29**, 122–133 (2008).

26. Pakotiprapha, D., Liu, Y., Verdine, G. L. & Jeruzalmi, D. A structural model for the damage-sensing complex in bacterial nucleotide excision repair. *J. Biol. Chem.* **284**, 12837–12844 (2009).
27. Pakotiprapha, D., Samuels, M., Shen, K. N., Hu, J. H. & Jeruzalmi, D. Structure and mechanism of the UvrA–UvrB DNA damage sensor. *Nat. Struct. Mol. Biol.* **19**, 291–U247 (2012).
28. Jaciuk, M. et al. A combined structural and biochemical approach reveals translocation and stalling of UvrB on the DNA lesion as a mechanism of damage verification in bacterial nucleotide excision repair. *DNA Repair* **85**, 102746 (2020).
29. Nguyen, B., Ordabayev, Y., Sokoloski, J. E., Weiland, E. & Lohman, T. M. Large domain movements upon UvrD dimerization and helicase activation. *Proc. Natl Acad. Sci. USA* **114**, 12178–12183 (2017).
30. Duchi, D., Mazumder, A., Malinen, A. M., Ebright, R. H. & Kapanidis, A. N. The RNA polymerase clamp interconverts dynamically among three states and is stabilized in a partly closed state by ppGpp. *Nucleic Acids Res.* **46**, 7284–7295 (2018).
31. Zuo, Y. H., Wang, Y. M. & Steitz, T. A. The mechanism of *E. coli* RNA polymerase regulation by ppGpp is suggested by the structure of their complex. *Mol. Cell* **50**, 430–436 (2013).
32. Kawale, A. A. & Burmann, B. M. UvrD helicase–RNA polymerase interactions are governed by UvrD’s carboxy-terminal Tudor domain. *Commun. Biol.* **3**, 607 (2020).
33. Sanders, K. et al. The structure and function of an RNA polymerase interaction domain in the PcrA/UvrD helicase. *Nucleic Acids Res.* **45**, 3875–3887 (2017).
34. Urrutia-Irazabal, I., Ault, J. R., Sobott, F., Savery, N. J. & Dillingham, M. S. Analysis of the PcrA–RNA polymerase complex reveals a helicase interaction motif and a role for PcrA/UvrD helicase in the suppression of R-loops. *eLife* **10**, e68829 (2021).

35. Manelyte, L. et al. The unstructured C-terminal extension of UvrD interacts with UvrB, but is dispensable for nucleotide excision repair. *DNA Repair* **8**, 1300–1310 (2009).
36. Martinez, B., Bharati, B. K., Epshtain, V. & Nudler, E. Pervasive transcription-coupled DNA repair in *E. coli*. *Nat. Commun.*, <https://doi.org/10.1038/s41467-022-28871-y> (2022).
37. Courcelle, J., Khodursky, A., Peter, B., Brown, P. O. & Hanawalt, P. C. Comparative gene expression profiles following UV exposure in wild-type and SOS-deficient *Escherichia coli*. *Genetics* **158**, 41–64 (2001).
38. Thomassen, G. O. et al. Tiling array analysis of UV treated *Escherichia coli* predicts novel differentially expressed small peptides. *PLoS ONE* **5**, e15356 (2010).
39. Lin, L. L. & Little, J. W. Autodigestion and RecA-dependent cleavage of Ind<sup>-</sup> mutant LexA proteins. *J. Mol. Biol.* **210**, 439–452 (1989).
40. Richardson, J. P. Preventing the synthesis of unused transcripts by Rho factor. *Cell* **64**, 1047–1049 (1991).
41. Jain, S., Gupta, R. & Sen, R. Rho-dependent transcription termination in bacteria recycles RNA polymerases stalled at DNA lesions. *Nat. Commun.* **10**, 1207 (2019).
42. Bryant, J. A., Sellars, L. E., Busby, S. J. W. & Lee, D. J. Chromosome position effects on gene expression in *Escherichia coli* K-12. *Nucleic Acids Res.* **42**, 11383–11392 (2014).
43. Perdiz, D. et al. Distribution and repair of bipyrimidine photoproducts in solar UV-irradiated mammalian cells. Possible role of Dewar photoproducts in solar mutagenesis. *J. Biol. Chem.* **275**, 26732–26742 (2000).
44. Ikenaga, M., Ichikawa-Ryo, H. & Kondo, S. The major cause of inactivation and mutation by 4-nitroquinoline 1-oixde in *Escherichia*

- coli*: excisable 4NQO-purine adducts. *J. Mol. Biol.* **92**, 341–356 (1975).
45. Zdravetski, Z. Z., Mello, J. A., Marinus, M. G. & Essigmann, J. M. Multiple pathways of recombination define cellular responses to cisplatin. *Chem. Biol.* **7**, 39–50 (2000).
  46. Wade, J. T. & Grainger, D. C. Pervasive transcription: illuminating the dark matter of bacterial transcriptomes. *Nat. Rev. Microbiol.* **12**, 647–653 (2014).
  47. Kunala, S. & Brash, D. E. Intragenic domains of strand-specific repair in *Escherichia coli*. *J. Mol. Biol.* **246**, 264–272 (1995).
  48. Gaul, L. & Svejstrup, J. Q. Transcription-coupled repair and the transcriptional response to UV-irradiation. *DNA Repair* **107**, 103208 (2021).
  49. Duan, M., Selvam, K., Wyrick, J. J. & Mao, P. Genome-wide role of Rad26 in promoting transcription-coupled nucleotide excision repair in yeast chromatin. *Proc. Natl Acad. Sci. USA* **117**, 18608–18616 (2020).
  50. Oh, J., Xu, J., Chong, J. & Wang, D. Molecular basis of transcriptional pausing, stalling, and transcription-coupled repair initiation. *Biochim. Biophys. Acta Gene Regul. Mech.* **1864**, 194659 (2021).
  51. Velez-Cruz, R. & Egly, J. M. Cockayne syndrome group B (CSB) protein: at the crossroads of transcriptional networks. *Mech. Ageing Dev.* **134**, 234–242 (2013).
  52. Ghosh-Roy, S., Das, D., Chowdhury, D., Smerdon, M. J. & Chaudhuri, R. N. Rad26, the transcription-coupled repair factor in yeast, is required for removal of stalled RNA polymerase-II following UV irradiation. *PLoS ONE* **8**, e72090 (2013).
  53. Li, W. T. & Li, S. S. Facilitators and repressors of transcription-coupled DNA repair in *Saccharomyces cerevisiae*. *Photochem. Photobiol.* **93**, 259–267 (2017).

54. Mayne, L. V. & Lehmann, A. R. Failure of RNA synthesis to recover after UV irradiation: an early defect in cells from individuals with Cockayne's syndrome and xeroderma pigmentosum. *Cancer Res.* **42**, 1473–1478 (1982).
55. Zhao, D. D. et al. CRISPR/Cas9-assisted gRNA-free one-step genome editing with no sequence limitations and improved targeting efficiency. *Sci Rep.* **7**, 16624 (2017).
56. Baba, T. et al. Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. *Mol. Syst. Biol.* **2**, 2006.0008 (2006).
57. Datsenko, K. A. & Wanner, B. L. One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proc. Natl Acad. Sci. USA* **97**, 6640–6645 (2000).
58. Kitagawa, M. et al. Complete set of ORF clones of *Escherichia coli* ASKA library (a complete set of *E. coli* K-12 ORF archive): unique resources for biological research. *DNA Res.* **12**, 291–299 (2005).
59. Schneider, C. A., Rasband, W. S. & Eliceiri, K. W. NIH Image to ImageJ: 25 years of image analysis. *Nat. Methods* **9**, 671–675 (2012).
60. Carter, J. D. & LaBean, T. H. Coupling strategies for the synthesis of peptide-oligonucleotide conjugates for patterned synthetic biomineralization. *J. Nucleic Acids* **2011**, 926595 (2011).
61. Ghosh, S. S., Kao, P. M., Mccue, A. W. & Chappelle, H. L. Use of maleimide-thiol coupling chemistry for efficient syntheses of oligonucleotide-enzyme conjugate hybridization probes. *Bioconjugate Chem.* **1**, 71–76 (1990).
62. Hermanson, G. T. *Bioconjugate Techniques* 3rd edn (Academic, 2013).
63. Grabarek, Z. & Gergely, J. Zero-length crosslinking procedure with the use of active esters. *Anal. Biochem.* **185**, 131–135 (1990).

64. Staros, J. V., Wright, R. W. & Swingle, D. M. Enhancement by *N*-hydroxysulfosuccinimide of water-soluble carbodiimide-mediated coupling reactions. *Anal. Biochem.* **156**, 220–222 (1986).
65. Lu, L. et al. Identification of MS-cleavable and noncleavable chemically cross-linked peptides with MetaMorpheus. *J. Proteome Res.* **17**, 2370–2376 (2018).
66. Wisniewski, J. R. Label-free and standard-free absolute quantitative proteomics using the “total protein” and “proteomic ruler” approaches. *Methods Enzymol.* **585**, 49–60 (2017).
67. Meyer, A. et al. Systematic analysis of protein-detergent complexes applying dynamic light scattering to optimize solutions for crystallization trials. *Acta Crystallogr. F* **71**, 75–81 (2015).
68. Chi, H. et al. Comprehensive identification of peptides in tandem mass spectra using an efficient open search engine. *Nat. Biotechnol.* **36**, 1059–1061 (2018).
69. Chen, Z. L. et al. A high-speed search engine pLink 2 with systematic evaluation for proteome-scale identification of cross-linked peptides. *Nat. Commun.* **10**, 3404 (2019).
70. Spivak, G. & Hanawalt, P. C. Determination of damage and repair in specific DNA Sequences. *Methods* **7**, 147–161 (1995).
71. Iyer, S., Park, B. R. & Kim, M. Absolute quantitative measurement of transcriptional kinetic parameters in vivo. *Nucleic Acids Res.* **44**, e142 (2016).
72. Dannenmann, B. et al. Simultaneous quantification of DNA damage and mitochondrial copy number by long-run DNA-damage quantification (LORD-Q). *Oncotarget* **8**, 112417–112425 (2017).
73. Lehle, S. et al. LORD-Q: a long-run real-time PCR-based DNA-damage quantification method for nuclear and mitochondrial genome analysis. *Nucleic Acids Res.* **42**, e41 (2014).

74. Rothfuss, O., Gasser, T. & Patenge, N. Analysis of differential DNA damage in the mitochondrial genome employing a semi-long run real-time PCR approach. *Nucleic Acids Res.* **38**, e24 (2010).
75. Zhu, S. & Coffman, J. A. Simple and fast quantification of DNA damage by real-time PCR, and its application to nuclear and mitochondrial DNA from multiple tissues of aging zebrafish. *BMC Res. Notes* **10**, 269 (2017).
76. Crowley, D. J. & Hanawalt, P. C. Induction of the SOS response increases the efficiency of global nucleotide excision repair of cyclobutane pyrimidine dimers, but not 6-4 photoproducts, in UV-irradiated *Escherichia coli*. *J. Bacteriol.* **180**, 3345–3352 (1998).
77. Koehler, D. R., Courcelle, J. & Hanawalt, P. C. Kinetics of pyrimidine(6-4)pyrimidone photoproduct repair in *Escherichia coli*. *J. Bacteriol.* **178**, 1347–1350 (1996).
78. Kai-Feng, H., Sidorova, J. M., Nghiem, P. & Kawasumi, M. The 6-4 photoproduct is the trigger of UV-induced replication blockage and ATR activation. *Proc. Natl Acad. Sci. USA* **117**, 12806–12816 (2020).
79. Davis, S. E. et al. Mapping *E. coli* RNA polymerase and associated transcription factors and identifying promoters genome-wide. *Methods Enzymol.* **498**, 449–471 (2011).
80. Lee, C., Kim, J., Shin, S. G. & Hwang, S. Absolute and relative QPCR quantification of plasmid copy number in *Escherichia coli*. *J. Biotechnol.* **123**, 273–280 (2006).
81. Epshteyn, V. & Nudler, E. Cooperation between RNA polymerase molecules in transcription elongation. *Science* **300**, 801–805 (2003).
82. Yang, J. & Zhang, Y. I-TASSER server: new development for protein structure and function predictions. *Nucleic Acids Res.* **43**, W174–W181 (2015).

83. Rossi, F. et al. The biological and structural characterization of *Mycobacterium tuberculosis* UvrA provides novel insights into its mechanism of action. *Nucleic Acids Res.* **39**, 7316–7328 (2011).
84. Vogel, U. & Jensen, K. F. NusA is required for ribosomal antitermination and for modulation of the transcription elongation rate of both antiterminated RNA and mRNA. *J. Biol. Chem.* **272**, 12265–12271 (1997).
85. Belogurov, G. A. & Artsimovitch, I. The mechanisms of substrate selection, catalysis, and translocation by the elongating RNA polymerase. *J. Mol. Biol.* **431**, 3975–4006 (2019).
86. Canutescu, A. A., Shelenkov, A. A. & Dunbrack, R. L. A graph-theory algorithm for rapid protein side-chain prediction. *Protein Sci.* **12**, 2001–2014 (2003).
87. Krieger, E. & Vriend, G. New ways to boost molecular dynamics simulations. *J. Comput. Chem.* **36**, 996–1007 (2015).
88. Orban-Nemeth, Z. et al. Structural prediction of protein models using distance restraints derived from cross-linking mass spectrometry data. *Nat. Protoc.* **13**, 1724–1724 (2018).
89. Duhovny, D., Nussinov, R. & Wolfson, H. J. In *Algorithms in Bioinformatics. WABI 2002. Lecture Notes in Computer Science* vol. 2452 (eds Guigó, R. & Gusfield, D.) 185–200 (Springer, 2002).
90. Schneidman-Duhovny, D., Inbar, Y., Nussinov, R. & Wolfson, H. J. PatchDock and SymmDock: servers for rigid and symmetric docking. *Nucleic Acids Res.* **33**, W363–W367 (2005).
91. Bullock, J. M. A., Schwab, J., Thalassinos, K. & Topf, M. The importance of non-accessible crosslinks and solvent accessible surface distance in modeling proteins with restraints from crosslinking mass spectrometry. *Mol. Cell. Proteomics* **15**, 2491–2500 (2016).

92. Martin, A. C. R. & Porter, C. T. ProFit V3.1  
<http://www.bioinf.org.uk/software/profit/> (2009).
93. Chen, V. B. et al. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D* **66**, 12–21 (2010).
94. Severinov, K., Mooney, R., Darst, S. A. & Landick, R. Tethering of the large subunits of *Escherichia coli* RNA polymerase. *J. Biol. Chem.* **272**, 24137–24140 (1997).
95. Zhang, Y. et al. Structural basis of transcription initiation. *Science* **338**, 1076–1080 (2012).
96. Opalka, N. et al. Complete structural model of *Escherichia coli* RNA polymerase from a hybrid approach. *PLoS Biol.* **8**, e1000483 (2010).
97. Skubak, P. et al. A new MR-SAD algorithm for the automatic building of protein models from low-resolution X-ray data and a poor starting model. *IUCrJ* **5**, 166–171 (2018).
98. Cowtan, K. The Buccaneer software for automated model building. 1. Tracing protein chains. *Acta Crystallogr. D* **62**, 1002–1011 (2006).
99. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D* **66**, 486–501 (2010).
100. Murshudov, G. N. et al. REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr. D* **67**, 355–367 (2011).
101. Nakagawa, N. et al. Crystal structure of *Thermus thermophilus* HB8 UvrB protein, a key enzyme of nucleotide excision repair. *J. Biochem.* **126**, 986–990 (1999).
102. Lee, S. J., Sung, R. J. & Verdine, G. L. Mechanism of DNA lesion homing and recognition by the Uvr nucleotide excision repair system. *Research* **2019**, 5641746 (2019).

103. Orren, D. K. & Sancar, A. The (A)BC excinuclease of *Escherichia coli* has only the UvrB and Uvrc subunits in the incision complex. *Proc. Natl Acad. Sci. USA* **86**, 5237–5241 (1989).
104. Zou, Y. & Van Houten, B. Strand opening by the UvrA(2)B complex allows dynamic recognition of DNA damage. *EMBO J.* **18**, 4889–4901 (1999).
105. Verhoeven, E. E. A., Wyman, C., Moolenaar, G. F. & Goosen, N. The presence of two UvrB subunits in the UvrAB complex ensures damage detection in both DNA strands. *EMBO J.* **21**, 4196–4205 (2002).
106. Brugger, C. et al. Molecular determinants for dsDNA translocation by the transcription-repair coupling and evolvability factor Mfd. *Nat. Commun.* **11**, 3740 (2020).
107. Deaconescu, A. M., Sevostyanova, A., Artsimovitch, I. & Grigorieff, N. Nucleotide excision repair (NER) machinery recruitment by the transcription-repair coupling factor involves unmasking of a conserved intramolecular interface. *Proc. Natl Acad. Sci. USA* **109**, 3353–3358 (2012).

## Acknowledgements

We thank the staff at beamline BL17U1/BL18U1/BL19U1 of Shanghai Synchrotron Radiation Facility for assistance during data collection, and T. Artemev for his continued support. The work was supported by the National Key Research and Development Program of China (grant no. 2018YFA0903701), the Strategic Priority Research Program of the Chinese Academy of Sciences (XDB29020302), the Chinese Natural Science Foundation of China (31822001) and the Shanghai Science and Technology Innovation program (19JC1415900) (Y.Z.), and by NIH grant R01 GM126891, the Blavatnik Family Foundation and the Howard Hughes Medical Institute (E.N.).

## Author information

## Author notes

1. These authors contributed equally: Binod K. Bharati, Manjunath Gowder

## Affiliations

1. Department of Biochemistry and Molecular Pharmacology, New York University School of Medicine, New York, NY, USA

Binod K. Bharati, Manjunath Gowder, Khaled Alzoubi, Vladimir Svetlov, Venu Kamarthapu, Jacob W. Weaver, Vitaly Epshtein, Nikita Vasilyev & Evgeny Nudler

2. Howard Hughes Medical Institute, New York University School of Medicine, New York, NY, USA

Binod K. Bharati, Manjunath Gowder, Venu Kamarthapu & Evgeny Nudler

3. Key Laboratory of Synthetic Biology, Chinese Academy of Sciences Center for Excellence in Molecular Plant Sciences, Institute of Plant Physiology and Ecology, Chinese Academy of Sciences, Shanghai, China

Fangfang Zheng, Liqiang Shen & Yu Zhang

## Contributions

B.K.B. generated bacterial strains, performed the pull-down experiments described in Fig. 1 and performed CPD repair and other experiments described in Fig. 4 and Extended Data Figs. 1, 4–6. M.G. generated bacterial strains and performed DNA repair and other experiments described in Fig. 5 and Extended Data Figs. 6–8, 10. F.Z. and L.S. purified and crystallized the proteins, determined the X-ray crystal structures and performed other experiments described in Fig. 3a and Extended Data Fig. 3. K.A. performed XLMS-driven structural docking and modelling. V.S. performed XLMS and DLS experiments. V.K. and J.W.W. purified the

proteins and reconstituted pre-TCRCs and TCRCs. V.K. and B.K.B. performed bacterial survival tests. V.E. performed in vitro transcription assays and UvrD biochemical analysis. N.V. performed mass-spectrometry analysis. Y.Z. supervised the structural work and assisted with data analysis. E.N. designed and supervised the project and wrote the manuscript with input from all authors. All authors discussed the results.

## Corresponding authors

Correspondence to [Yu Zhang](#) or [Evgeny Nudler](#).

## Ethics declarations

## Competing interests

The authors declare no competing interests.

## Peer review

## Peer review information

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

## Additional information

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## Extended data figures and tables

[\*\*Extended Data Fig. 1 Quantification of UvrABD binding to RNAP in vivo.\*\*](#)

Supplementary to Fig. 1. **a**, UV sensitivity of *E. coli* strains used in Fig. 1. Representative efficiencies of colony formation of parent wild-type (MG1655) and mutant *E. coli* cells exposed to the indicated UV doses. Overnight cultures were diluted 1:100 with fresh LB and grown to  $\sim 2 \times 10^7$ . Cells were serially diluted, plated on LB agar, irradiated with UV, and incubated at 37 °C for 24 h. **b**, Representative western blots used to generate the plots of Fig. 1b–d. **c**, Quantitative mass-spectrometry analysis of RNAP-associated UvrABD, Mfd, and Rho in the exponentially grown cells prior to genotoxic stress. RNAP pulldown samples were prepared as in Fig. 1. Values are normalized to that of NusA-containing RNAPs, thus reflecting the RNAP molecules engaged in elongation in vivo.

## Extended Data Fig. 2 XLMS-driven structural modelling of TCRC.

Supplementary to Fig. 2. **a–d**, Reconstitution of the TCRC without nucleic acids. **a**, Isolation of the RNAP–NusA–UvrABD complex by SEC. SDS-Coomassie gel represents the protein fraction eluted from the main peak (P). **b**, DLS analysis of the RNAP–NusA–UvrABD complex. The ‘P’ fraction from **a** was subjected to DLS. Raleigh sphere (R) estimate of the complex molecular weight (MW = 908 kDa), which deviates by only 1.7% from the theoretical MW of a uniform monodispersed complex containing 1 RNAP, 1 NusA, 2 UvrD, 2 UvrA, and 1 UvrB molecules. **c**, Network view of the highly confident non-redundant inter-protein cross-links (Supplementary Table 1). **d**, XLMS-based model of the reconstituted RNAP:NusA:UvrABD complex. The model was built based on the in vitro cross-links using PatchDock and the workflow described in Extended Data Fig. 11 and Methods. **e–h**, Mapping UvrD-EC interactions in vitro. **e**, Isolation of the EC:NusA:UvrD complex by SEC. SDS-Coomassie gel represents the protein fraction eluted from the main SEC peak (P). **f**, DLS analysis of the EC:NusA:UvrD complex. “P” fraction from (e) was subjected to DLS. Raleigh sphere (R) estimate of the complex molecular weight (MW = 620 kDa), which deviates by only 2% from the theoretical MW of a uniform monodispersed complex containing 1 RNAP, 1 NusA, and 2 UvrD molecules. **g**, Network view of the highly confident non-redundant inter-protein cross-links between RNAP subunits, NusA, and UvrD (Supplementary Table 1). **h**, XLMS-based model of the

RNAP:NusA:UvrD complex (Extended Data Fig. 11 and Methods). The model shows the positioning of UvrD monomers relative to the transcription bubble. Blue star indicates the DNA-UvrD cross-link previously mapped in the EC<sup>18</sup>. CTD of UvrD2 is shown by green hexagon. RNA is not shown.

### Extended Data Fig. 3 Structural interrogation of UvrD CTD interactions in TCRC.

Supplementary to Fig. 3a. **a–g**, Structural analysis of UvrD CTD–RNAP interactions. **a**, UvrD and RNAP β-subunit domains observed in crystals structure were coloured and labelled. **b**, The overall structure of the UvrD CTD–RNAP β2i4 (PDB: 7EGS, Extended Data Table 1) complex. The major interface is highlighted by a rectangle. The ‘N’ and ‘C’ termini of UvrD CTD are numbered. **c**, Detailed UvrD CTD–RNAP β2i4 interactions. Oxygen, nitrogen, and water atoms are coloured in red, blue, and orange, respectively. Blue dash, H-bond. **d**, Yeast two-hybrid assay results show that alanine substitution of interface residues on UvrD CTD or βi4 impairs the interaction of UvrD and RNAP β pincer. The potential interactions were selected on SD (-HALW) plates, and the growth on SD (-LW) plates was used as input control. **e**, Strep-tag pull down results show that alanine substitution of interface residues of UvrD CTD or βi4 impairs RNAP–UvrD interaction. **f**, Sequence alignments of UvrD CTD and RNAP β2i4 from 316 non-redundant proteobacteria that contain βi4 insertion on RNAP. Key interface residues were labelled with asterisks and numbered as in *E. coli*. **g**, Cys pair cross-linking results demonstrate direct proximity of UvrD CTD and RNAP βi4. Wild-type or mutated UvrD-RNAP complexes were incubated in oxidative ( $\text{CuCl}_2$ ) or reducing (DTT) condition and separated by SDS-PAGE. The asterisk marks two major impurity bands. **h–m**, Structural analysis of UvrD CTD–UvrB interactions. **h**, UvrD and UvrB domains (numbered as in *E. coli*). The domains observed in crystal structure are highlighted in colours. **i**, UvrD CTD interacts with UvrB-1a/1b/2domain (or UvrD NTD), consistent with a previous report<sup>35</sup>. *E. coli* UvrD and UvrB were fused to GAL4-AD and GAL4-DBD, respectively. The potential interactions were selected on SD (-HALW) plates, and the growth on SD (-LW) plates was used as input control. **j**, A 2.6-Å crystal structure of *T.*

*thermophilus* UvrD CTD–UvrB NTD complex (PDB: 7EGT, Extended Data Table 1). UvrB-1a docks on a shallow groove of UvrD CTD. UvrD CTD, UvrB-1a, UvrB-1b, and UvrB-2 are coloured in purple, cyan, orange, and light green, respectively. **k**, Detailed UvrD CTD–UvrB-1a interactions. Residues H654, R656, K680, R681, S683 of UvrD CTD make a H-bond network with D27, E29, R30, Q383 of UvrB-1a (residues are numbered as in *T. thermophilus*; the corresponding residues in *E. coli* are indicated in parentheses). Y686 of UvrD CTD makes stacking interaction with R55 and Q383 of UvrB-1a. **l**, Structural superimposition of UvrB NTD/UvrD CTD complex (coloured as above) and UvrB/dsDNA complex (grey and red; PDB: 6O8F)<sup>102</sup> shows that UvrD CTD binds the opposite surface of UvrB dsDNA-loading cleft, implicating UvrD doesn't affect dsDNA loading Of UvrB. **m**, Structural comparison between UvrD CTD/UvrB NTD (left) and UvrD CTD/RNAP  $\beta$ 2i4 (right, PDB: 7EGS) shows that UvrB and RNAP  $\beta$ 2i4 binds at the same cleft of UvrD CTD, and thereby suggests that the interactions of UvrB and RNAP to UvrD are mutually exclusive.

#### **Extended Data Fig. 4 Functional analysis of UvrD( $\Delta$ CTD) and RNAP( $\Delta$ $\beta$ i4).**

Supplementary to Fig. 3. **a**, Deletions of the CTD of UvrD or  $\beta$ i4 of RNAP partially compromise UvrD-mediated backtracking. EC20 was formed by the wild-type RNAP or RNAP lacking  $\beta$ i4 (green  $\Delta$ ) (lanes 13 to 18) at the T7A1 DNA template and then chased in the presence of specified amounts of UvrD (red  $\Delta$ ). The pro-backtracking activity of UvrD was assessed as a ratio (%) between the amount of full length (run-off) product and total amounts of RNA products located below the run-off. Majority of these products are the result of UvrD-mediated backtracking and sensitive to transcript cleavage by GreB<sup>12,18</sup>. **b**, Deletion of the CTD does not compromise UvrD catalytic activity. The autoradiogram shows the thin layer chromatography (TLC) plate of UvrD-mediated ATP hydrolysis. The reaction was performed using polyC single stranded DNA template as described in Methods. The means  $\pm$  SE from three experiments are plotted on the right. **c**, *uvrD* $^{\Delta}$ CTD and  $\Delta$  $\beta$ i4 cells are equally more sensitive to genotoxic stress as compared to wild type. Representative efficiencies of colony formation of wild-type (MG1655) and mutant cells following treatment with the indicated dose of UV irradiation. Cells were grown to

$\text{OD}_{600\text{ nm}} \sim 0.4$  and serial 10-fold dilutions were spotted on LB agar plates followed by UV irradiation and incubation in the dark at 37 °C for 24 h.

### **Extended Data Fig. 5 NER strictly coupled to transcription and is mostly independent of Mfd.**

Supplementary to Fig. 4. **a**, Wild-type (wt) and mutant cells were UV irradiated at 50 J/m<sup>2</sup> and allowed to recover. At the indicated times, genomic DNA was isolated and either treated with T4endoV or mock treated for 30 min at 37 °C and then analysed on alkali-agarose gels. Rifampicin (Rif; 750 µg/ml), chloramphenicol (Cm; 200 µg/ml) and/or bicyclomycin (Bcm; 25 µg/ml) were added 30 min prior to UV irradiation (see [Methods](#)). Representative gels are shown for each analysed strain and condition. The percentage of repaired (lesion-free) DNA in T4endoV-treated samples is plotted for each time point relative to untreated samples (see Fig. 4). Data are the mean ± SEM from at least three independent experiments. **b**, Mfd overexpression interferes with global NER. The wild-type cells containing pMfd and the “empty” vector control (pCA24N) were induced with 0.1 mM IPTG followed by UV irradiation and recovery. The percentage of repaired (lesion-free) DNA in T4endoV-treated samples is plotted for each time point relative to untreated samples. Data are the mean ± SEM from at least three independent experiments.

### **Extended Data Fig. 6 High Rif abolishes NER.**

Supplementary to Fig. 4. **a, b**, Effects of high (750 µg/ml) and low (50 µg/ml) Rif on *E. coli* transcription and NER. **a**, Inhibition of chromosomal *lacZ* transcription by Rif. Copies/µL cDNA of *lacZ* transcripts was determined using absolute quantification (see [Methods](#)). A standard curve was generated using *lacZ* PCR product ( $10^{16}$  to  $10^{23}$ ). RT-qPCR was performed using cDNA isolated from bacterial cultures treated with indicated concentrations of Rif. Number of copies of *lacZ* transcripts was determined by interpolation. Values are the mean ± SD from 3 independent experiments. **b**, Inhibition of CPDs repair by high and low Rif. (Left panel) Representative slot blot probed by fluorescently labelled secondary antibody to reveal binding of primary monoclonal CPD-specific antibodies

(see [Methods](#)). (Right panel) Quantitative analysis of slot blot images for the indicated recovery time points post-UV. Bars, standard errors of the means from 3 independent experiments. **c, d**, Rif does not compromise the level of NER enzymes during the time of the experiment. **c**, Representative western blots of intracellular UvrABCD proteins during the time of high Rif treatment (Fig. 4) and their quantification (**d**). Data are the mean ± SEM from at least three independent experiments.

## [Extended Data Fig. 7 Local transcription enables NER.](#)

Supplementary to Fig. 5. **a–d**, Depriving the genomic loci of transcription abolishes their NER. **a**, Schematics of the *mCherry* insulators. A transcription unit containing *mCherry* (with or without *lacZ* promoter) flanked by the terminator cassettes was inserted at the *tam* and *nupG* loci. **b–d** The expression of *mCherry* from the insulators (**b**) and RNAP occupancy (**c, d**) after IPTG induction, as determined by RT-qPCR and ChIP–qPCR, respectively. Values are the mean ± SD from 3 independent experiments. **e, f**, CPD repair within the insulators. Most of NER strictly depends on promoter-initiated transcription. The levels of transcription and NER are stronger within the *nupG* insulator comparing to the *tam* insulator. Cells were induced with IPTG followed by UV irradiation (40 J/m<sup>2</sup>) and recovery in the dark for the indicated time intervals. CPD density was determined by SLR-qPCR as in Fig. 5a and used to calculate the percentage of repaired CPDs. Values are the mean ± SD from 3 independent experiments. **g–j**, UvrAB recruitment to the UV-damaged DNA strictly depends on local transcription in both *tam* and *nupG* loci. Occupancy of RNAP (**c, d**), UvrA (**g, i**) and UvrB (**h, j**) following UV irradiation was determined by ChIP–qPCR. Cells were induced with IPTG followed by UV irradiation (40 J/m<sup>2</sup>) and recovery for the indicated time intervals. Values are the mean ± SD from 3 independent experiments. Results are shown as a fold change in the occupancy of UvrAB within the insulator following UV irradiation. UT – untreated. Values are mean ± SD from 3 independent experiments. \*\*P < 0.01, \*\*\*P < 0.0001 (Statistical analysis was performed using unpaired non-parametric two-tailed Mann-Whitney t test). P values compare the percentage of DNA repair between “promoter” and “no promoter” strains for a given time point.

## Extended Data Fig. 8 Local transcription enables NER irrespective of Mfd.

Supplementary to Fig. 5. **a–h**, Depriving the genomic loci of transcription drastically diminished NER irrespective of Mfd. Genomic DNA repair within the insulator was monitored as described in Methods for the lesions generated by 4-NQO (**a, e**), NFZ (**b, f**), cisplatin (**c, g**), or UV-C (**d, h**). *lacZ* was induced with IPTG followed by the exposure to drugs or UV radiation. Cells were allowed to recover for the indicated time intervals followed by the isolation of genomic DNA. Lesion density was determined by SLR-qPCR and used to calculate the percentage of repaired lesions. Values are the mean  $\pm$  SD from 3 independent experiments. **i, j**, UvrAB recruitment to the UV-damaged DNA strictly depends on local transcription, but not Mfd (compare to Fig. 5l, m). Recruitment of UvrA (**i**) and UvrB (**j**) to the *lacZ* insulator (with or without promoter) was determined by ChIP-qPCR in  $\Delta mfd$  cells as in Fig. 5h. Results are shown as a fold change in the occupancy of UvrAB within the insulator of  $\Delta mfd$  following UV irradiation. \*\*P < 0.01, \*\*\*\*P < 0.0001. Statistical analysis was performed using unpaired non-parametric two-tailed Mann-Whitney t test. Values are the mean  $\pm$  SD from 3 independent experiments.

## Extended Data Fig. 9 Integrated model of bacterial NER.

See Supplementary Videos 2, 3. **a–f**, Based on the in vivo and in vitro data presented, we propose a structure-functional model of NER in *E. coli* in which the elongating RNAP functions as the primary lesion scanner and platform for the assembly of active NER complexes. **a**, A subpopulation of elongating RNAPs persistently interacts with UvrD1 and UvrA1, as shown in Fig. 2b. The in vivo RNAP pulldowns and XLMS demonstrate that such surveillance pre-TCRCs can form even before the genotoxic stress. **b, c**, On stalling at a lesion (CPD is marked as “TT”), pre-TCRC recruits UvrD2 to form a helicase competent UvrD dimer. UvrD2 CTD (green hexagon) interacts with RNAP  $\beta$ i4 to stabilize the UvrD dimer. UvrD12 pulls TCRC backward, thereby exposing a CPD to NER enzymes<sup>18</sup>. ppGpp contributes at this stage by rendering RNAP backtracking-prone<sup>12</sup>. **d**, TCRC recruits UvrA2/UvrB to initiate the lesion processing. Although a single UvrB

monomer is sufficient for lesion verification and UvrC recruitment<sup>28,103,104</sup>, the second UvrB molecule may be recruited as well<sup>27,102,105</sup>. In vitro (Extended Data Fig. 2a–d) and in vivo XLMS (Fig. 2c, d) are consistent with a single UvrB monomer model. This UvrB can interact with the CTD of UvrD2 (Extended Data Fig. 3h–m), thereby displacing UvrD2 from RNAP (Supplementary Video 2). The release of UvrD2 that occurs coincidentally with the UvrA2B recruitment (Fig. 1b, c) supports such a sequence of events in vivo. UvrD2 displacement would abrogate any further UvrD-mediated backtracking. e, The pre-incision TCRC recruits UvrC and releases UvrA2B followed by the NER execution step<sup>5,103</sup>. f, Once repair has been completed, the backtracked pre-TCRC is promptly recovered by the anti-backtracking factors (GreB, Mfd, and a leading ribosome) to resume elongation. g–k, Role of Mfd in NER (Supplementary Video 3). We propose that the modest contribution of Mfd to NER (Fig. 4a) is due to its ability to terminate multiple queuing ECs in front of TCRCs, thereby helping to “clean up” space between pre-TCRC/TCRCs and DNA lesions at highly expressed genes. g, h, UvrA of pre-TCRC facilitates Mfd recruitment and/or its transition to a processive translocase (Fig. 1d). i, Mfd then translocates forward (downstream of TCRC) to terminate multiple ECs between the TCRC and CPD (red “TT”). This directionality ensures that Mfd preferentially terminates non-TCR complexes, thereby facilitating TCRC access to damage sites. j, TCR proceeds as in (a–f). k, Mfd continues to be recruited even after most repair has been completed (Fig. 1d). These additional Mfd molecules can now also reactivate backtracked complexes, hence the role of Mfd in facilitating transcription recovery post-UV<sup>14</sup>. This model explains why a delay in NER in Mfd-deficient cells occurs only within the most highly transcribed (most congested) regions and why TCR of less actively transcribed regions is indifferent to, or even compromised by, Mfd activity (for example, Fig. 5c, d)<sup>36</sup>. It also explains why the overexpression of Mfd is so detrimental to NER<sup>36</sup> (Extended Data Fig. 5b): excessive Mfd would prematurely terminate both ECs and TCRCs, thereby abolishing repair. Finally, the model also explains why *mfd* cells become more sensitive to genotoxic stress in the presence of Rho inhibitor Bcm (Extended Data Fig. 10). Rho, like Mfd, can terminate ECs that obscure the lesion sites from TCRCs<sup>41</sup>. If both termination factors were inactivated, there is no obvious solution to this problem. l, Structural model

for UvrA-mediated Mfd recruitment. (top) *E. coli* Mfd (colour: cyan) bound to DNA (PDB: 6XEO)<sup>106</sup> is shown interacting with UvrA in pre-TCRC by structural alignment to the *E. coli* UvrA-Mfd complex (PDB: 4DFC)<sup>107</sup>. UvrA-binding domain of Mfd (colour: blue) is fully unmasked after initial DNA binding, allowing it to be located by pre-TCRC (**g–k** and Supplementary Video 3). RNAP-binding domain of Mfd (green) is facing downstream to eventually interact with RNAPs stalled/paused ahead of the pre-TCRC and terminate or rescue them from the backtracked state. Illustrative cartoon (bottom). RNA is not shown.

### **Extended Data Fig. 10 Rho inactivation enhances Mfd sensitivity to UV.**

**a**, Representative efficiencies of colony formation of wild-type (MG1655) and  $\Delta mfd$  cells on LB agar, LB agar exposed to 40 J/m<sup>2</sup> of UV, 25  $\mu$ M bicyclomycin (Bcm) and 25  $\mu$ M Bcm with 40 J/m<sup>2</sup> of UV. Cells were spotted on LB agar plates in serial tenfold dilutions and incubated at 37 °C for 24 h. **b**, Data from three independent experiments was used to calculate percent difference in survival between wild-type and  $\Delta mfd$  cells. Values are means  $\pm$  SD from 3 independent experiments.

### **Extended Data Fig. 11 Modelling of pre-TCRC/TCRC with restraints from XLMS.**

**a**, An outline of the automated workflow for cross-link-guided docking. **i**, Coordinate files for all the *E. coli* interactors used to build the pre-TCRC and TCRC are prepared using available PDB structures, which were refined using YASARA Structure<sup>87</sup> (see [Methods](#)). Proteins without available PDB structures were homology-modelled using I-TASSER<sup>82,88</sup>. XLMS data was converted to the distance restraints compatible with PatchDock. **ii**, PDB files of receptor and ligand molecules were submitted to PatchDock with their corresponding distance restraints for rigid-body docking. **iv**, The docking results were validated by examining the cross-link satisfaction using Jwalk<sup>91</sup> (see [Methods](#)). **v**, The docked complexes were ranked by the number/fraction of cross-links validated by Jwalk and the average solvent accessible surface distance (SASD). **b**, An overview of the automated

dimer-assembly workflow. Two monomer models (X and Y), previously and separately docked to a common receptor model (R), are combined to generate receptor-dimer models that satisfy the highest number of cross-links between the two monomers. **i**, **ii**, Top docking poses for each monomer (R-X and R-Y) are clustered to eliminate redundancies and accelerate subsequent steps. **iii**, Representative models obtained by clustering each of the two groups are cross-matched to generate combined receptor-dimer coordinate files (R-X-Y), and analysed for cross-links satisfied between X-Y using Jwalk. **iv**, receptor-dimer models satisfying  $> 2$  cross-links are ranked by number and average distance of satisfied X-Y cross-links for further analysis.

### **Extended Data Fig. 12 Application of the docking pipeline.**

**a–d**, pre-TCRC modelling. **a**, PDB files of *E. coli* ECs were downloaded and prepared by extracting chains corresponding to RNAP subunits and NusA, then refined using YASARA Structure. *E. coli* UvrA was modelled using the homology template server I-TASSER. **b**, UvrA was docked to ECs using PatchDock, with the RNAP-UvrA cross-links as distance restraints. **c**, PDB coordinates file of *E. coli* UvrD in the apo form was trimmed to the first 640 residues and refined, then docked to the top EC-UvrA (as ranked by cross-link satisfaction) from the previous step. **d**, EC-UvrA-UvrD complexes generated in the previous step were ranked by RNAP–UvrD cross-link satisfaction and used as receptors to dock UvrD CTD. Results were clustered using ProFit (v.3.1), and finally analysed for the UvrAD DNA-binding regions alignment with the DNA path in the EC. **e–g**, TCRC modelling. **e**, Same as **a**. Docking was repeated using EC-UvrA1 complexes as receptors and additional UvrA-UvrA distance restraints to generate EC-UvrA12 complexes. **f**, Same as **(c)**. Structure then docked to the top EC-UvrA12 complexes generated in the previous step, as ranked by RNAP-UvrA and UvrA-UvrA cross-link satisfaction. **g**, Top EC-UvrA12-UvrD complexes generated in the previous step were divided into two groups based on the docked UvrD model (apo vs. closed), and used as input to the dimer-assembly component (Extended Data Fig. [11](#) and Methods). UvrD poses from the two groups were cross-matched to generate EC-UvrA12-UvrD12 complexes, analysed for UvrD-UvrD cross-link satisfaction and steric clashes, and used as receptors to dock UvrD1 CTD.

Results were clustered using ProFit (v.3.1) and analysed for agreement with UvrA-dimer structures and alignment of Uvr DNA-binding regions with the DNA path in the EC. Final complexes were refined with YASARA Structure and re-analysed with Jwalk for cross-link satisfaction.

## **Extended Data Table 1 Statistics of X-ray crystal structures**

## **Supplementary information**

### **Supplementary Information**

This file contains Supplementary Discussion, Supplementary Figs. 1–5, Supplementary Tables 1–3, Supplementary References and legends for Supplementary Videos 1–3.

### **Reporting Summary**

### **Supplementary Video 1**

XLMS-driven structural model of TCRC.

### **Supplementary Video 2**

Integrated model of NER in *E. coli*.

### **Supplementary Video 3**

Mfd functioning in NER.

## **Rights and permissions**

### **Reprints and Permissions**

## **About this article**

## Cite this article

Bharati, B.K., Gowder, M., Zheng, F. *et al.* Crucial role and mechanism of transcription-coupled DNA repair in bacteria. *Nature* **604**, 152–159 (2022). <https://doi.org/10.1038/s41586-022-04530-6>

- Received: 11 April 2021
- Accepted: 07 February 2022
- Published: 30 March 2022
- Issue Date: 07 April 2022
- DOI: <https://doi.org/10.1038/s41586-022-04530-6>

## Share this article

Anyone you share the following link with will be able to read this content:

[Get shareable link](#)

Sorry, a shareable link is not currently available for this article.

[Copy to clipboard](#)

Provided by the Springer Nature SharedIt content-sharing initiative

---

This article was downloaded by **calibre** from <https://www.nature.com/articles/s41586-022-04530-6>

- Article
- [Published: 30 March 2022](#)

# Targeting *Xist* with compounds that disrupt RNA structure and X inactivation

- [Rodrigo Aguilar](#) [ORCID: orcid.org/0000-0003-0410-5478](#)<sup>1,2</sup> nAff<sup>5</sup>,
- [Kerrie B. Spencer](#)<sup>3</sup>,
- [Barry Kesner](#)<sup>1,2</sup>,
- [Noreen F. Rizvi](#)<sup>3</sup>,
- [Maulik D. Badmalia](#)<sup>4</sup>,
- [Tyler Mrozowich](#)<sup>4</sup>,
- [Jonathan D. Mortison](#)<sup>3</sup>,
- [Carlos Rivera](#)<sup>1,2</sup>,
- [Graham F. Smith](#)<sup>3</sup>,
- [Julja Burchard](#)<sup>3</sup>,
- [Peter J. Dandliker](#)<sup>3</sup>,
- [Trushar R. Patel](#) [ORCID: orcid.org/0000-0003-0627-2923](#)<sup>4</sup>,
- [Elliott B. Nickbarg](#)<sup>3</sup> &
- [Jeannie T. Lee](#) [ORCID: orcid.org/0000-0001-7786-8850](#)<sup>1,2</sup>

[Nature](#) volume 604, pages 160–166 (2022)

- 7550 Accesses
- 183 Altmetric
- [Metrics details](#)

## Subjects

- [Chromosomes](#)
- [Long non-coding RNAs](#)
- [Nucleic-acid therapeutics](#)
- [Pharmaceutics](#)

## Abstract

Although more than 98% of the human genome is non-coding<sup>1</sup>, nearly all of the drugs on the market target one of about 700 disease-related proteins. The historical reluctance to invest in non-coding RNA stems partly from requirements for drug targets to adopt a single stable conformation<sup>2</sup>. Most RNAs can adopt several conformations of similar stabilities. RNA structures also remain challenging to determine<sup>3</sup>. Nonetheless, an increasing number of diseases are now being attributed to non-coding RNA<sup>4</sup> and the ability to target them would vastly expand the chemical space for drug development. Here we devise a screening strategy and identify small molecules that bind the non-coding RNA prototype *Xist*<sup>5</sup>. The X1 compound has drug-like properties and binds specifically the RepA motif<sup>6</sup> of *Xist* in vitro and in vivo. Small-angle X-ray scattering analysis reveals that RepA can adopt multiple conformations but favours one structure in solution. X1 binding reduces the conformational space of RepA, displaces cognate interacting protein factors (PRC2 and SPEN), suppresses histone H3K27 trimethylation, and blocks initiation of X-chromosome inactivation. X1 inhibits cell differentiation and growth in a female-specific manner. Thus, RNA can be systematically targeted by drug-like compounds that disrupt RNA structure and epigenetic function.

This is a preview of subscription content

## Access options

Subscribe to Nature+

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

\$29.99

monthly

[Subscribe](#)

Subscribe to Journal

Get full journal access for 1 year

\$199.00

only \$3.90 per issue

[Subscribe](#)

All prices are NET prices.

VAT will be added later in the checkout.

Tax calculation will be finalised during checkout.

Buy article

Get time limited or full article access on ReadCube.

\$32.00

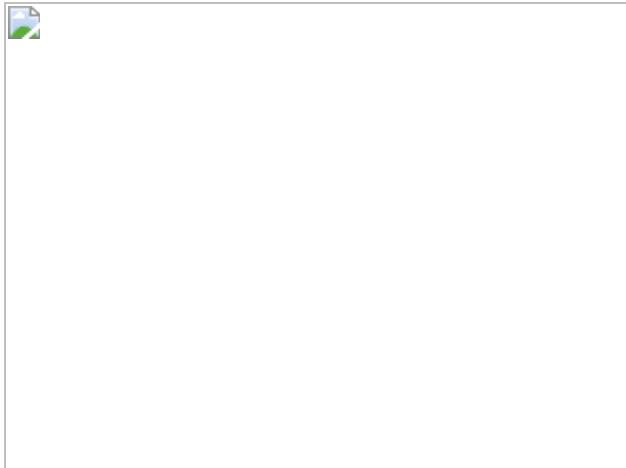
[Buy](#)

All prices are NET prices.

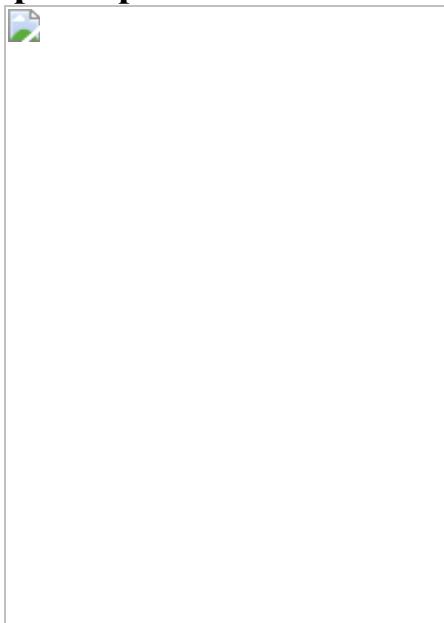
**Additional access options:**

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

**Fig. 1: X1 binds RepA and weakens RNA binding to interacting protein partners in vitro.**



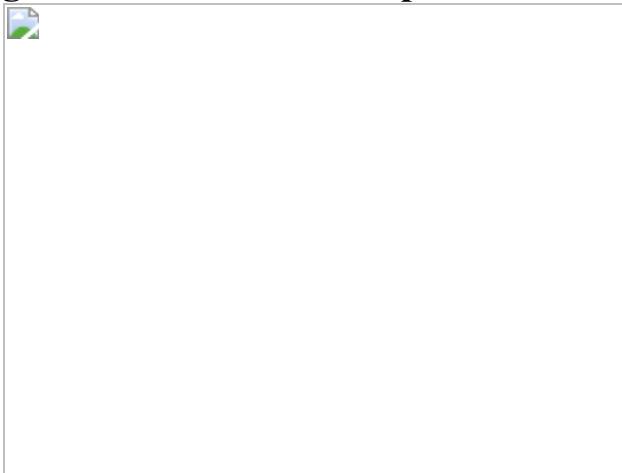
**Fig. 2: X1 weakens the *Xist* interaction with PRC2 and SPEN in a RepA-dependent manner inside cells.**



**Fig. 3: X1 treatment leads to Xi-specific loss of PRC2 and H3K27me3 enrichment and failure of XCI.**



**Fig. 4: 3D structure of RepA with and without X1.**



## Data availability

Sequence data that support the findings of this study have been deposited in the Gene Expression Omnibus (GEO) with accession number [GSE141683](#).

## References

1. Consortium, E. P. et al. Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature* **447**, 799–816 (2007).

2. Santos, R. et al. A comprehensive map of molecular drug targets. *Nat. Rev. Drug Discov.* **16**, 19–34 (2017).
3. Warner, K. D. et al. Principles for targeting RNA with drug-like small molecules. *Nat. Rev. Drug Discov.* **17**, 547–558 (2018).
4. Zhang, F. & Lupski, J. R. Non-coding genetic variants in human disease. *Hum. Mol. Genet.* **24**, R102–R110 (2015).
5. Lee, J. T. Epigenetic regulation by long noncoding RNAs. *Science* **338**, 1435–1439 (2012).
6. Wutz, A. et al. Chromosomal silencing and localization are mediated by different domains of *Xist* RNA. *Nat. Genet.* **30**, 167–174 (2002).
7. Rizvi, N. F. & Smith, G. F. RNA as a small molecule druggable target. *Bioorg. Med. Chem. Lett.* **27**, 5083–5088 (2017).
8. Disney, M. D. et al. Drugging the RNA world. *Cold Spring Harb. Perspect. Biol.* **10**, a034769 (2018).
9. Howe, J. A. et al. Selective small-molecule inhibition of an RNA structural element. *Nature* **526**, 672–677 (2015).
10. Palacino, J. et al. SMN2 splice modulators enhance U1-pre-mRNA association and rescue SMA mice. *Nat. Chem. Biol.* **11**, 511–517 (2015).
11. Disney, M. D. et al. Informa 2.0: a platform for the sequence-based design of small molecules targeting structured RNAs. *ACS Chem. Biol.* **11**, 1720–1728 (2016).
12. Allen Annis, D. et al. An affinity selection-mass spectrometry method for the identification of small molecule ligands from self-encoded combinatorial libraries—Discovery of a novel antagonist of E-coli dihydrofolate reductase. *Int. J. Mass Spectrom.* **238**, 77–83 (2004).
13. Rizvi, N. F. et al. Discovery of selective RNA-binding small molecules by affinity-selection mass spectrometry. *ACS Chem. Biol.* **13**, 820–831

(2018).

14. Rizvi, N. F. et al. Targeting RNA with small molecules: identification of selective, RNA-binding small molecules occupying drug-like chemical space. *SLAS Discov.* **25**, 384–396 (2020).
15. Cifuentes-Rojas, C. et al. Regulatory interactions between RNA and polycomb repressive complex 2. *Mol. Cell* **55**, 171–185 (2014).
16. Chillon, I. et al. Native purification and analysis of long RNAs. *Methods Enzymol.* **558**, 3–37 (2015).
17. Lipinski, C. A. Lead- and drug-like compounds: the rule-of-five revolution. *Drug Discov. Today Technol.* **1**, 337–341 (2004).
18. Monfort, A. et al. Identification of Spen as a crucial factor for *Xist* function through forward genetic screening in haploid embryonic stem cells. *Cell Rep.* **12**, 554–561 (2015).
19. Lee, M. K. et al. A novel small-molecule binds to the influenza A virus RNA promoter and inhibits viral replication. *Chem. Commun.* **50**, 368–370 (2014).
20. Ogawa, Y. et al. Intersection of the RNA interference and X-inactivation pathways. *Science* **320**, 1336–1341 (2008).
21. Zhao, J. et al. Genome-wide identification of polycomb-associated RNAs by RIP-seq. *Mol. Cell* **40**, 939–953 (2010).
22. Patil, D. P. et al. m<sup>6</sup>A RNA methylation promotes *XIST*-mediated transcriptional repression. *Nature* **537**, 369–373 (2016).
23. Sunwoo, H. et al. Repeat E anchors *Xist* RNA to the inactive X chromosomal compartment through CDKN1A-interacting protein (CIZ1). *Proc. Natl Acad. Sci. USA* **114**, 10654–10659 (2017).
24. Jeon, Y. & Lee, J. T. YY1 tethers *Xist* RNA to the inactive X nucleation center. *Cell* **146**, 119–133 (2011).

25. Colognori, D. et al. *Xist* deletional analysis reveals an interdependency between *Xist* RNA and polycomb complexes for spreading along the inactive X. *Mol. Cell* **74**, 101–117 e110 (2019).
26. Liu, F. et al. Visualizing the secondary and tertiary architectural domains of lncRNA RepA. *Nat. Chem. Biol.* **13**, 282–289 (2017).
27. Kikhney, A. G. & Svergun, D. I. A practical guide to small angle X-ray scattering (SAXS) of flexible and intrinsically disordered proteins. *FEBS Lett.* **589**, 2570–2577 (2015).
28. Kim, D. N. et al. Zinc-finger protein CNBP alters the 3-D structure of lncRNA *Braveheart* in solution. *Nat. Commun.* **11**, 148 (2020).
29. Carrette, L. L. G. et al. A mixed modality approach towards Xi reactivation for Rett syndrome and other X-linked disorders. *Proc. Natl Acad. Sci. USA* **115**, 1715124115 (2017).
30. Stelzer, A. C. et al. Discovery of selective bioactive small molecules by targeting an RNA dynamic ensemble. *Nat. Chem. Biol.* **7**, 553–559 (2011).

## Acknowledgements

We thank all members of the Lee and Patel laboratories, Charles Lesberg and other Merck team members for stimulating scientific discussions. We acknowledge the Merck MINT Award, a grant from the US National Institutes of Health (R01-HD097665), and funding from the Howard Hughes Medical Institute to J.T.L.; the Pew Charitable Trust Latin American Fellows Program to R.A. and C.R.; and the MGH Fund for Medical Discovery to R.A. M.D.B. and T.M. acknowledge MITACS and NSERC PGS-D fellowships respectively. T.R.P. is Canada Research Chair in RNA and Protein Biophysics. We thank DIAMOND Light Source B21 beamline staff for their help with data collection (BAG SM22113).

## Author information

## Author notes

### 1. Rodrigo Aguilar

Present address: Institute of Biomedical Sciences (ICB), Faculty of Medicine and Faculty of Life Sciences, Universidad Andres Bello, Santiago, Chile

## Affiliations

### 1. Department of Molecular Biology, Massachusetts General Hospital, Boston, MA, USA

Rodrigo Aguilar, Barry Kesner, Carlos Rivera & Jeannie T. Lee

### 2. Department of Genetics, The Blavatnik Institute, Harvard Medical School, Boston, MA, USA

Rodrigo Aguilar, Barry Kesner, Carlos Rivera & Jeannie T. Lee

### 3. Merck & Co., Boston, MA, USA

Kerrie B. Spencer, Noreen F. Rizvi, Jonathan D. Mortison, Graham F. Smith, Julja Burchard, Peter J. Dandliker & Elliott B. Nickbarg

### 4. Alberta RNA Research and Training Institute, Department of Chemistry and Biochemistry, University of Lethbridge, Lethbridge, Alberta, Canada

Maulik D. Badmalia, Tyler Mrozowich & Trushar R. Patel

## Contributions

R.A., K.S., B.K., N.R., T.M., M.D.B., J.M., G.S., J.B., P.J.D., T.R.P., E.N. and J.T.L. designed experiments and interpreted data. R.A. performed RNA purifications, biochemical assays, immunoFISH, ChIP-seq and RNA-seq. K.S. coordinated the small molecule synthesis and off-target activity evaluations. N.R. and E.N. performed the ALIS determinations. C.R.

performed RNA purifications. B.K. conducted bioinformatics analyses. T.M., M.D.B. and T.R.P. performed and analysed SAXS data. R.A., K.S. and J.T.L. wrote the paper.

## Corresponding author

Correspondence to [Jeannie T. Lee](#).

## Ethics declarations

## Competing interests

K.B.P., N.F.R., J.D.M., G.F.S., J.B., P.J.D. and E.B.N. are current or former employees of Merck & Co. and may hold stock or other financial interests in Merck & Co. J.T.L. is a cofounder of Translate Bio and Fulcrum Therapeutics and is also a scientific advisor to Skyhawk Therapeutics.

## Peer review

## Peer review information

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

## Additional information

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## Extended data figures and tables

### [Extended Data Fig. 1 Purification of Xist RepA RNA.](#)

A 431 Repeat A fragment of Xist RNA was *in vitro* transcribed and purified under native conditions by FPLC. A representative chromatogram is shown.

To confirm size and stability of the sample just prior to ALIS, we visualized the RNA in a denaturing urea-PAGE.

**Extended Data Fig. 2 X1 inhibits interaction of Xist RepA with cognate interacting proteins *in vitro*.**

**a**, RNA EMSAs show that X1 weakens interaction between RepA and PRC2, and RepA and SPEN-RRM. Increasing concentrations of the compounds (0, 5, 7.5, 10, 25, 50, 75, 100  $\mu$ M) were titrated against 0.5 nM RNA and 15.6 nM PRC2, or 0.1 nM RNA and 158 nM SPEN-RRM. Two replicates showed similar results. **b**, RNA EMSAs titrating PRC2 (0, 15.6, 31.2, 62.4, 124.9, 250 nM) or SPEN-RRM (0, 79.2, 158, 396, 792, 1580 nM) against a fixed concentration of X1 (25 or 75  $\mu$ M) and 0.5 nM RepA, Tsix (reverse complement of RepA), or Hotair RNA—all of which are known PRC2 interactors. For SPEN, RNA was 0.1 nM. Two or more replicates showed similar results. **c**, Increasing concentrations of X1 (0, 5, 7.5, 10, 25, 50, 75, 100  $\mu$ M) was titrated against 0.5 nM RNA (Tsix, Hotair) and 15.6 nM PRC2. One representative gel of two replicates is shown. **d**, Densitometric analysis to determine IC<sub>50</sub>, which were too high to be measured for Tsix and Hotair. Data are represented as mean +/– SD. n = 2 independent experiments. RepA result from Fig. [1f](#) is shown as reference. **e**, Order of addition does not affect X activity. Increasing concentrations of the compounds (0, 5, 7.5, 10, 25, 50, 75, 100  $\mu$ M) was titrated against 0.5 nM RepA and 15.6 nM PRC2. One representative gel of two replicates is shown. **Top**, PRC2 was added to a RepA-molecule pre-incubated mix. **Bottom**, Molecule was added to a RepA-PRC2 pre-incubated mix.

**Extended Data Fig. 3 X1 also inhibits interaction of Xist RepA with cognate interacting proteins *in vivo*.**

**a-b**, RIP-qPCR analysis in d4 female TST ES cells to evaluate Xist binding to EZH2 (**a**) and RBM15 (**b**) in 10  $\mu$ M X1. IgG, negative control antibody. Other EZH2 interactors Malat1, Gtl2, Htr6-us and Nespas are shown. Gapdh, negative control RNA. Bars: mean. Individual data points included. n = 2 biologically independent experiments quantified in duplicate (Representative graph shown). **c**, **Top**: RT-qPCR confirms similar quantities

of Xist RNA in control and X-A samples prior to Xist RNA pulldown. Xist exons 4-5 primers were used. **Bottom:** Similar quantities were also present following Xist RNA pulldown, thereby ruling out unequal Xist expression as a cause of unequal H<sup>3</sup> radioactive counts. X-A cells amplified poorly with RepA primers, consistent with deletion of RepA.

**Extended Data Fig. 4 X1 effects on EB outgrowth in ♀ - TST-XX, ♀ -XO, and ♂ -XY EB cells.**

**a**, Growth of differentiating ♀ -TST cells at day 3, or 24 h post-X1 treatment, up to 10 μM X1. Data are represented as Tukey box plots. Lower whisker: 25<sup>th</sup> percentile minus 1.5xInterquartile Range (IQR). Higher whisker: 75<sup>th</sup> percentile plus 1.5xIQR. Box range: 25<sup>th</sup> (bottom) to 75<sup>th</sup> (top) percentile. Line within box: median. Points beyond higher whisker are shown. *P*-values: one-way ANOVA with respect to control cells. n = 150 colonies combined from 3 independent experiments. **b**, Viability of d5 cells. n = 3 biologically independent experiments. **c**, No obvious effect on day 3 female EB growth after 24 h X1 treatment. **d**, Quantitation of EB outgrowth at day 5 (72 h post-drug application). The distance from EB center to edge of outgrowth was measured in 100 d3 or 30 d5 EBs combined from 3 independent experiments. Data presented as in panel (a). *P*-values: one-way ANOVA with respect to control cells. **e**, Weaker effect of X16 on ♀ -TST EB outgrowth. No obvious effect of X-negative. One representative brightfield microscopy from 3 independent cultures is shown. Center of the EB and edge of outgrowth as marked. Scale as indicated. **f**, X1 had no effect on growth of pre-XCI (d0) female cells. **g,h**, X1 also did not inhibit ♀ -TST-XO and ♂ -XY ES cells at day 3 (**g**) or day 5 (**h**). Neither cell line expresses Xist or undergoes XCI. One representative field is shown. Scale bar, 150 μm. **i**, Quantitation of EB outgrowth in XY male and XO female EBs at days 3 and 5. Distance from the EB center to the edge of outgrowth was measured. Day 3: n = 136, XO colonies; n = 112, XY colonies. Day 5: n = 40, XO colonies; n = 60, XY colonies). Data presented as in panel (a).

**Extended Data Fig. 5 Karyotype analysis of ES cells and RNA immunoFISH analysis of day 3 X1-treated cells.**

**a**, X-chromosome painting DNA FISH of DMSO- and X1-treated XX TST cells, and a DMSO-treated XO clone that spontaneously arose from the XX TST cells. Scale as shown. Inset: magnification of representative nucleus. %nuclei with indicated X chromosome number shown. n, sample size combining from 3 biologically independent experiments. **b**, Xist/Tsix RNA-FISH and immunostaining for H3K27me3, H2AK119ub, EZH2, and RING1B in ♀ -TST EB at day 3. One representative nucleus is shown. %cells with Xist foci is indicated. n, sample size. Scale bar, 5 μm.

**Extended Data Fig. 6 Full fields for RNA immunoFISH experiments of Fig. 2 and Extended Data Fig. 5.**

**a**, Full fields for the RNA FISH and Immunofluorescence experiments, with boxed nuclei presented in Fig. 2g and Extended Data Fig. 5b. **b**, Full fields for H3K27me3 immunostaining of DMSO- or X1-treated ♀ TST-A cells, with boxed nuclei presented in Fig. 2h. %cells with foci on the Xi as indicated (sample size, n, from two biologically independent experiments combined). **c**, Western blot using H3K27me3 and total histone H3 antibodies. Total cell extracts were obtained from day 7 female EB cells after treating with 10 μM of various compounds from day 2. Compounds: EZH2 inhibitor 1 (EPZ-6438, MedChem Express), EZH2 inhibitor 2 (PF-06821497, Pfizer), or X1. One representative film of two replicates is shown.

**Extended Data Fig. 7 Epigenomic analyses of PRC2 and H3K27me3 enrichment.**

**a–b**, Allele-specific H3K27me3 (**a**) and SUZ12 (**b**) ChIP-seq analyses of day 5 female EB treated with 10 μM X1 or DMSO (control) for 72 h. Tracks for all reads (composite, "comp"), mus (Xi), and cas (Xa). Dotted green lines separate ChrX. **c–f**, Zoom-ins for allele-specific H3K27me3 ChIP-seq analyses of day 5 female EB treated with 10 μM X1 or DMSO (control) for 72 h. Browser shots shown with sliding window 1 kb, step size 0.5 kb. Scale shown in brackets. **c**, X-linked genes subjected to XCI. **d**, the *Xist* gene. **e**, Escapees. **f**, Representative control autosomal gene on Chr13. **g**, Box plot of normalized read densities for the –5000 to +1 region of ChrX

and Chr13 refSeq genes, parsed into mus and cas alleles. Lower whisker: 10<sup>th</sup> percentile. Higher whisker: 90<sup>th</sup> percentile. Box range: 25<sup>th</sup> (bottom) to 75<sup>th</sup> (top) percentile. Line within box: median. Points beyond whiskers are shown. *P*-values: two-tailed Wilcoxon test from data gathered from individual H3K27me3 and Suz12 ChIP-seq experiments.

### **Extended Data Fig. 8 Analysis of gene expression and X1 reversibility.**

**a**, Time course RT-qPCR of indicated control genes in DMSO- or X1-treated female EB. X1 added on indicated days (pink arrows). Mean and S.D. shown for 3 biological replicates. **b**, Time course allele-specific RT-qPCR of indicated Xa genes in DMSO- or X1-treated female EB. X1 added on indicated days (pink arrows). Mean and S.D. shown for 3 biological replicates. **c**, Dose-response analysis in the range of 0–10 µM X1 compound. Allele-specific RT-qPCR of indicated X-linked genes in DMSO- or X1-treated female EB. Mus allele (Xi) shown. X1 was added on d2. *P*, two-tailed Student's *t*-test with respect to DMSO-treated TST control. Mean and S.D. shown for 2 replicates. At 10 µM X1, the Student's *t*-test reveal no significant difference between d7 cells and expression found in control ES cells. **d–e**, Female EB were grown from d1 in 10 µM X1 and the treatment was suspended on day 3, 4, 5, or maintained up to day 7. The growth morphology (**d**) and Mecp2 expression from the Xi is evaluated at d7 (**e**). One representative brightfield microscopy from 3 independent cultures is shown.

### **Extended Data Fig. 9 Transcriptomic studies of on- and off-target effects.**

**a–d**, RNA-seq analyses of day 5 DMSO- or X1-treated female EB. Zoom-ins to representative X-linked genes subjected to XCI (**a**), Xist (**b**), escapee gene (**c**) and autosomal gene (**d**). Tracks for all reads (comp), mus reads (Xi), and cas reads (Xa). FPM scale shown in brackets. **e**, Differentially expressed autosomal genes (y axis) and their corresponding changes in H3K27me3 enrichment (x axis). Each dot represents a gene. Number of genes on each of the nine sections as shown. Comp tracks were sampled to

the smallest library, then MultiTesting and IndependentFiltering DESeq2 filtering was performed reporting significance below 0.05 (Wald test) after Benjamini and Hochberg correction with the application of independent intensity filtering.

### **Extended Data Fig. 10 Sixteen conformational clusters identified for native RepA RNA without X1 treatment.**

**a**, HPLC-SEC profile of the purified RepA RNA with or without X1 previous to SAXS data collection. **b**, PRIMUS analysis for initial data quality analysis. Inset: Guinier plot to determine the Radius of Gyration ( $R_g$ ). **c**, Dimensionless Kratky analysis [ $q^*Rg$  vs.  $I(q)/I(0)x(q^*Rg)^2$ ] of samples. **d**, Pairwise distance distribution profile ( $P(r)$ ) to estimate the real space dimensions of the molecule in Å. **e**, 16 clusters (C1–C16) of RepA are presented in their native state without X1. C13 is the dominant conformation. Pie-chart shows relative abundance of structural clusters. See also Fig. 4.

## **Supplementary information**

### **Supplementary Information**

Methods.

### **Reporting Summary**

### **Supplementary Figure 1**

Original source images for EMSAs and western blots.

### **Supplementary Table 1**

RNAs tested in ALIS surveys. The 42 transcripts tested against X22 or X1 are listed. The mouse *Xist* RepA RNA (the only X22 binder) is highlighted in yellow. RNAs with >60% GC content are highlighted in pink.

## **Supplementary Table 2**

Analogue compounds from Fig. 1c that demonstrated binding in affinity ranking experiments.

## **Supplementary Table 3**

Cellular uptake and concentration of X1 molecule by up to 180-fold. Here we addressed whether cells concentrated the X1 molecule following uptake. In brief, we added a known amount of tritiated X1 molecule to embryonic stem (ES) cells, differentiation day 7 embryoid bodies (EB), or fibroblasts (T4) (row 1). After 24 h, radioactivity was measured as counts per minute (cpm) from a total cell lysate using a scintillation counter (row 2). Knowing the cellular volume and the number of total cells in each assay (rows 3 and 4), we calculated the total volume of cells (row 5) into which X1 molecule was diluted ( $\text{cpm } \mu\text{l}^{-1}$ ) (row 6). The  $\text{cpm } \mu\text{l}^{-1}$  value was interpolated to a standard curve (sheet 2) constructed with increasing concentrations of the tritiated molecule:  $[\text{cpm } \mu\text{l}^{-1}] = 33.954 \times [\text{nM molecule}] - 898.7$ . The concentration factor could then be derived by comparing the concentration of the molecule found in cells (row 7) with the original amount added to the culture (row 1).

## **Supplementary Table 4**

Further demonstration of selectivity of X1 for *Xist* and X inactivation. List of autosomal genes where relative changes in RNA expression are  $>2$  or  $<-0.5$ , and relative changes in H3K27me3 enrichment are  $>4$  or  $<-0.25$ . All values are logarithmic (base 2).

## **Supplementary Table 5**

List of upregulated genes on the Xi following X1 treatment. Left, all active genes encoded in the X chromosome with FKPM  $> 0.5$  are listed. Right, genes encoded on  $X^{\text{mus}}$ , up-regulated with X1. Results are expressed as the relative change (%) in the fraction of reads transcribed from the mus allele with respect to the overall (mus+cas) expression.

## Supplementary Video 1

C13 (dominant) conformation of RepA RNA.

## Supplementary Video 2

C6' conformation of RepA RNA with X1 molecule.

## Rights and permissions

### Reprints and Permissions

## About this article

### Cite this article

Aguilar, R., Spencer, K.B., Kesner, B. *et al.* Targeting *Xist* with compounds that disrupt RNA structure and X inactivation. *Nature* **604**, 160–166 (2022). <https://doi.org/10.1038/s41586-022-04537-z>

- Received: 27 November 2019
- Accepted: 08 February 2022
- Published: 30 March 2022
- Issue Date: 07 April 2022
- DOI: <https://doi.org/10.1038/s41586-022-04537-z>

## Share this article

Anyone you share the following link with will be able to read this content:

[Get shareable link](#)

Sorry, a shareable link is not currently available for this article.

[Copy to clipboard](#)

Provided by the Springer Nature SharedIt content-sharing initiative

---

This article was downloaded by **calibre** from <https://www.nature.com/articles/s41586-022-04537-z>

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

[Download PDF](#)

- Article
- Open Access
- [Published: 30 March 2022](#)

# Rixosomal RNA degradation contributes to silencing of Polycomb target genes

- [Haining Zhou](#)<sup>1</sup>,
- [Chad B. Stein](#) [ORCID: orcid.org/0000-0003-4711-3611](#)<sup>2</sup>,
- [Tiasha A. Shafiq](#)<sup>1</sup>,
- [Gergana Shipkovenska](#)<sup>1</sup>,
- [Marian Kalocsay](#) [ORCID: orcid.org/0000-0002-4187-5829](#)<sup>3</sup>,
- [Joao A. Paulo](#) [ORCID: orcid.org/0000-0002-4291-413X](#)<sup>3</sup>,
- [Jiuchun Zhang](#)<sup>4</sup>,
- [Zhenhua Luo](#)<sup>5</sup>,
- [Steven P. Gygi](#) [ORCID: orcid.org/0000-0001-7626-0034](#)<sup>3</sup>,
- [Karen Adelman](#) [ORCID: orcid.org/0000-0001-5364-334X](#)<sup>2</sup> &
- [Danesh Moazed](#) [ORCID: orcid.org/0000-0003-0321-6221](#)<sup>1</sup>

[Nature](#) volume 604, pages 167–174 (2022)

- 17k Accesses
- 1 Citations
- 68 Altmetric
- [Metrics details](#)

## Subjects

- [Gene silencing](#)
- [Histone post-translational modifications](#)
- [Protein–protein interaction networks](#)
- [RNA decay](#)

## Abstract

Polycomb repressive complexes 1 and 2 (PRC1 and PRC2) are histone-modifying and -binding complexes that mediate the formation of facultative heterochromatin and are required for silencing of developmental genes and maintenance of cell fate<sup>1,2,3</sup>. Multiple pathways of RNA decay work together to establish and maintain heterochromatin in fission yeast, including a recently identified role for a conserved RNA-degradation complex known as the rixosome or RIX1 complex<sup>4,5,6</sup>. Whether RNA degradation also has a role in the stability of mammalian heterochromatin remains unknown. Here we show that the rixosome contributes to silencing of many Polycomb targets in human cells. The rixosome associates with human PRC complexes and is enriched at promoters of Polycomb target genes. Depletion of either the rixosome or Polycomb results in accumulation of paused and elongating RNA polymerase at Polycomb target genes. We identify point mutations in the RING1B subunit of PRC1 that disrupt the interaction between PRC1 and the rixosome and result in diminished silencing, suggesting that direct recruitment of the rixosome to chromatin is required for silencing. Finally, we show that the RNA endonuclease and kinase activities of the rixosome and the downstream XRN2 exoribonuclease, which degrades RNAs with 5' monophosphate groups generated by the rixosome, are required for silencing. Our findings suggest that rixosomal degradation of nascent RNA is conserved from fission yeast to human, with a primary role in RNA degradation at facultative heterochromatin in human cells.

## Main

The Polycomb group proteins have central roles in silencing of cell type-specific and growth-related control genes and their loss is associated with developmental abnormalities and cancer<sup>1,2,3</sup>. Two major Polycomb complexes with histone-modifying and -binding activities have been identified. In the canonical PRC1 (cPRC1) complex, the RING1A or RING1B protein associate with PCGF2 or PCGF4, PHC, and chromobox (CBX) proteins<sup>7</sup>. RING1A and RING1B are RING finger E3 ubiquitin ligases that form the catalytic core of PRC1 complexes and mediate the mono-ubiquitination of histone H2A lysine 119<sup>8,9</sup> (H2AK119ub1). The PRC2 complex, consisting of EED, SUZ12, RBBP4 or RBBP7, and the EZH1 or EZH2 methyltransferases, methylates histone H3 lysine 27<sup>10,11,12,13</sup> (H3K27). In addition to cPRC1, variant PRC1 (vPRC1) complexes, which lack CBX proteins and along with RING1A or B contain PCGF1, 3, 5 or 6 and RYBP or YAF2<sup>14</sup>, have been identified. Each PRC complex can recognize the modification it catalyses as well as the one catalysed by the other complex. Thus H3K27 trimethylation (H3K27me3) is recognized by the EED subunit of PRC2 itself and the CBX subunit of cPRC1s<sup>15</sup>, and H2AK119ub1 is recognized by RYBP–vPRC1 and PRC2 accessory subunits JARID2

and AEBP2<sup>16,17,18</sup>. This crosstalk creates reinforcing positive-feedback loops that may promote the epigenetic inheritance of silencing<sup>2,3</sup>.

H2AK119ub1 has a key role in initiating the cascade of modifications that lead to the formation of Polycomb domains. The direct recruitment of vPRC1 to DNA and subsequent ubiquitination of H2AK119 lead to the recruitment of PRC2, deposition of H3K27me3 and cPRC1 binding<sup>19,20,21</sup>. Notably, the RING1B and CBX2 subunits of PRC1 can mediate chromatin compaction in vitro and in vivo<sup>22,23</sup>, and CBX2 in particular can mediate liquid–liquid phase separation<sup>24,25,26,27,28,29</sup>. The mechanism of silencing has therefore been proposed to involve the exclusion of RNA polymerase II (PolII) via compaction or condensate formation<sup>1,24,25,26,27,28,29</sup>. However, in mouse embryonic stem cells, vPRC1 complexes lacking chromatin compaction activity contribute to silencing largely independently of cPRC1<sup>30,31</sup>. Previous studies also provide evidence for the presence of the general transcription machinery and PolII at promoters of Polycomb-repressed genes<sup>1,32,33</sup>, suggesting that mechanisms beyond chromatin compaction contribute to Polycomb silencing.

The rixosome is a highly conserved and essential multienzyme complex whose major role is in ribosomal RNA (rRNA) processing and ribosome biogenesis<sup>6</sup>. It contains an endonuclease subunit (human LAS1L), which cleaves within the rRNA internal transcribed spacer 2 and generates a precursor with a 5'-OH group. The polynucleotide kinase subunit (human NOL9) of the complex then phosphorylates the precursor in a step that is required for XRN2-mediated trimming and the generation of mature 26S rRNA. In the fission yeast *Schizosaccharomyces pombe*, the rixosome associates with heterochromatin and is required for the spreading of histone H3 lysine 9 (H3K9) methylation into actively transcribed regions and epigenetic inheritance of heterochromatin<sup>4,34</sup>. To test whether the human rixosome has similar roles in heterochromatin regulation, we purified the complex from human cells and analysed its composition by mass spectrometry. We found that the rixosome associates with the human PRC1 and PRC2 complexes and is recruited to Polycomb target genes, where it promotes degradation of nascent RNA and release of PolII.

## Rixosome association with Polycomb

We used CRISPR–Cas9 genome editing in human embryonic kidney (HEK 293FT) cells to modify the endogenous copies of rixosome genes *NOL9* and *WDR18* to express 3×Flag–NOL9 and 3×Flag–WDR18 (Fig. 1a, Extended Data Fig. 1a). NOL9 and WDR18 are the human orthologues of the Grc3 and Crb3 subunits of the fission yeast rixosome, mutations of which disrupt heterochromatin maintenance<sup>4</sup>. As the rixosome also has an essential role in rRNA processing and ribosome biogenesis in nucleoli<sup>6</sup>, we used a fractionation protocol to enrich for chromatin-bound, rather than

nucleolar, rixosomes (Fig. 1b, c). We immunopurified 3×Flag–NOL9 and 3×Flag–WDR18 proteins (Fig. 1d, Extended Data Fig. 1b) and performed tandem-mass-tag mass spectrometry analysis of immunoprecipitates, which identified all seven known subunits of the rixosome—NOL9, WDR18, LAS1L, MDN1, PELP1, TEX10 and SENP3—in 3×Flag–NOL9 purifications (Fig. 1e, Extended Data Fig. 1a, Supplementary Table 7). In addition, subunits of PRC1 (RING1B), vPRC1.6 (L3MBTL2), PRC2 (EZH2, EED and SUZ12), and the PRC1-interacting ubiquitin protease (USP7) co-purified with 3×Flag–NOL9, but at lower efficiency than core rixosome components (Fig. 1e). Similarly, 3×Flag–WDR18 immunoprecipitations contained core rixosome, PRC1 (RING1B, RYBP and USP7), and PRC2 (EZH2 and RBBP4) subunits (Extended Data Fig. 1c). Consistent with these results, a previous study also identified rixosome subunits in immunoprecipitations of two different vPRC1 subunits<sup>35</sup> and a proteome-wide immunoprecipitation and mass spectrometry study found an association between CBX4 and multiple rixosome subunits<sup>36</sup>. In addition, both rixosome and PRC1 subunits were identified in purifications of CHTOP, a human chromatin-associated protein<sup>37</sup>. In contrast to fission yeast<sup>4,34</sup>, H3K9me-associated HP1 $\alpha$ , HP1 $\beta$  and HP1 $\gamma$ , were not significantly enriched in human rixosome purifications (Fig. 1e, Extended Data Fig. 1c), suggesting that in human cells the rixosome associates with heterochromatin modified with H2AK119ub1 and/or H3K27me3, rather than H3K9me3. We examined the above associations using immunoprecipitation and western blotting and found that (1) RING1B and EZH2 co-immunoprecipitated with Flag–NOL9 (Fig. 1f); (2) PELP1, NOL9, SENP3 and WDR18 rixosome subunits co-immunoprecipitated with RING1B (Fig. 1g); and (3) RING1B, but not SUV39H1, co-immunoprecipitated with several subunits of the rixosome (Extended Data Fig. 1d). These rixosome–Polycomb interactions were not sensitive to treatment with benzonase, suggesting that they occurred independently of RNA or DNA (Fig. 1f,g). In addition, mass spectrometry analysis of endogenously tagged and immunopurified PHC2–Flag and Flag–CBX4 showed that in addition to all of the subunits of cPRC1, the immunoprecipitates were enriched for the MDN1, WDR18 and PELP1 subunits of the rixosome (Fig. 1h, Extended Data Fig. 1e). These results demonstrate that the rixosome interacts physically with Polycomb complexes.

**Fig. 1: The rixosome interacts with the PRC1 and PRC2 complexes.**

---

 **figure 1**

**a**, Genomic tagging of endogenous *NOL9* with 3×Flag. **b**, Experimental design for protein immunoprecipitation from a chromatin fraction. **c**, Western blots showing fractionation of HEK 293FT cells. **d**, Silver-stained gel of Flag immunoprecipitations from wild-type and Flag–NOL9-expressing HEK 293FT cells. **e**, Volcano plot displaying results of tandem-mass-tag mass spectrometry of proteins enriched in Flag immunoprecipitations from Flag–NOL9-expressing cells relative to untagged cells from two independent experiments. *P* values calculated by two-sided *t*-test. Subunits

of the rixosome, PRC1, PCR2 and H3K9me3-associated HP1 proteins are highlighted. **f**, Immunoprecipitations (IP) showing the association of RING1B and EZH2 with Flag–NOL9 in HEK 293FT cells with or without benzonase treatment. IB, immunoblot. **g**, Immunoprecipitations showing the association of RING1B with rixosome subunits PELP1, TEX10, SENP3 and WDR18 in HEK 293FT cells with or without benzonase treatment. **h**, Volcano plot displaying mass spectrometry results of proteins enriched in Flag immunoprecipitations from cells expressing Flag–PHC2 relative to untagged cells from two independent experiments. *P* values calculated by two-sided *t*-test. Subunits of the rixosome, PRC1, PCR2 and selected other proteins are highlighted.

We next carried out yeast two-hybrid (Y2H) assays to identify potential direct rixosome–Polycomb interactions. These assays suggested that the rixosome subunit TEX10 interacts with CBX7, CBX8, EED and RING1B, and that PELP1 interacts with PCGF3 (Extended Data Fig. [1f](#)). Consistent with the immunoprecipitation–mass spectrometry results, we observed no interactions between HP1 proteins and any of the rixosome subunits that we tested (Extended Data Fig. [1f](#)). Therefore, in support of the biochemical data, the Y2H assays demonstrate interactions between the rixosome and PRC subunits.

As both immunoprecipitation–mass spectrometry and immunoprecipitation–western blotting identified RING1B as a rixosome-associated protein, and RING1B interacts with TEX10 in Y2H assays, we tested whether bacterially expressed and purified glutathione *S*-transferase (GST)–RING1B and TEX10 interacted in a pull-down assay. As shown in Extended Data Fig. [1g](#) and Extended Data Fig. [1h](#), lane 7, full-length GST–RING1B, but not GST alone, pulled down TEX10. This interaction was greatly diminished upon deletion of amino acids 121–140 in the coiled-coil domain 1 (CC1) of RING1B but was not affected by several other RING1B deletions (Extended Data Fig. [1h](#), compare lane 5 with other lanes; summarized in Extended Data Fig. [1g](#)). Furthermore, several amino acid substitutions within this domain, which did not affect RING1B expression, abolished the interaction of GST–RING1B with TEX10 (Extended Data Fig. [1i](#)). Together, these results identify RING1B–TEX10 as a direct contact point between the rixosome and PRC1.

## Rixosome and Polycomb co-localization

To examine the genome-wide localization of the rixosome in human cells, we carried out chromatin immunoprecipitation followed by high-throughput sequencing (ChIP–seq) in HEK 293FT cells using antibodies that recognize the TEX10 and MDN1 subunits of the rixosome. To control for antibody specificity, we performed ChIP–seq on cells treated with either control, TEX10-specific or MDN1-specific small interfering RNA (siRNA) (Extended Data Fig. [2a,b](#)). Correlation analysis with ChIP–

seq signals for Polycomb-catalysed histone modifications and RING1B showed highly correlated TEX10 ( $r = 0.85$ ) and MDN1 ( $r = 0.67$ ) colocalization with H2AK119ub1, high correlation for TEX10 colocalization with H3K27me3 ( $r = 0.43$ ), and very high correlation between TEX10 (0.86) and MDN1 (0.72) colocalization with RING1B (Fig. 2a). For comparison, the correlation between H2AK119ub1 and H3K27me3 ( $r = 0.43$ ) in these datasets was in a similar range (Fig. 2a). Consistently, heatmap analysis at all annotated transcription start sites (TSSs) showed similar enrichment patterns for TEX10, MDN1, RING1B, H2AK119ub1 and H3K27me3, but not H3K9me3 or H3K36me3, when we rank ordered genes by their TEX10 signal (Fig. 2b). TEX10-occupied genes also tended to exhibit H3K4me3, suggesting the presence of the rixosome at loci with engaged PolII, including bivalent Polycomb domains (containing both H3K27me3 and H3K4me3) (Fig. 2b). We then rigorously defined a set of TEX10-bound genes (Fig. 2c,  $n = 7,827$ ) and compared with similarly active TEX10-unbound genes ( $n = 13,177$ ), as described in Methods. We observed significant enrichment of H2AK119ub1 and H3K27me3 at TEX10-bound versus TEX10-unbound genes (Fig. 2d,e). By contrast, TEX10-bound genes were depleted of H3K36me3, whereas the TEX10-bound and unbound genes displayed similar enrichment for H3K4me3 (Extended Data Fig. 2c,d). Furthermore, TEX10 and MDN1 were enriched at TSSs (Extended Data Fig. 2e), as has been previously described for H2AK119ub1 and H3K27me3<sup>38</sup> (Extended Data Fig. 2f). When we repeated the co-occupancy analysis using peak calling, rather than enrichment relative to TSSs, we found that both RING1B and H2AK119ub1, but not H3K9me3, were enriched at TEX10- and MDN1-bound genomic regions (Extended Data Fig. 2g-j). At the single-gene level, genome browser snapshots of TEX10 and MDN1 ChIP-seq reads at the *PCDH10* gene showed co-enrichment with H2AK119ub1 and H3K27me3, but not H3K9me3, whereas the HOXA cluster was enriched for the rixosome subunits and H2AK119ub1 (Fig. 2f).

**Fig. 2: The rixosome localizes to transcription start sites with high PRC1 and PRC2 occupancy.**

---

 **figure 2**

**a**, Matrix depicting Spearman correlation coefficients between ChIP-seq datasets in HEK 293FT cells, calculated using summed read counts  $\pm 2$  kb from all annotated gene TSSs (hg19). **b**, Heatmap representations of ChIP-seq of TEX10, MDN1, RING1B and histone modifications (H2AK119ub1, H3K4me3, H3K27me3, H3K9me3,

---

H3K36me3, H3K79me3 and H3K4me3). Rank order is from highest to lowest TEX10 signal.  $\log_2$  enrichment was normalized to reads per genome coverage. Read counts per gene were averaged in 50-nucleotide (nt) bins. **c–e**, Average distribution of TEX10 (**c**), H2AK119ub1 (**d**) and H3K27me3 (**e**) ChIP-seq reads at TEX10-bound genes ( $n = 7,827$ ) versus TEX10-unbound genes ( $n = 13,177$ ). Read counts per gene were summed in 50-nt bins. All three factors are significantly enriched at TEX10-bound genes,  $P < 2.2 \times 10^{-16}$  for TEX10 (**c**),  $P < 2.2 \times 10^{-16}$  for H2AK119ub1 (**d**),  $P < 2.2 \times 10^{-16}$  for H3K27me3 (**e**); two-tailed Mann–Whitney test, using summed reads in the window  $\pm 2$  kb from TSS. See Extended Data Fig. 2c, d for other modifications. **f**, Genomic snapshots of ChIP-seq reads for the indicated experiments at the Polycomb target *PCDH10* gene and HOXA cluster.  $\log_2$  enrichment levels were normalized to reads per genome coverage.

Polycomb proteins localize to distinct foci in the nucleus, referred to as Polycomb bodies<sup>39,40</sup>. We next performed immunofluorescence staining using an antibody that recognizes the EZH2 subunit of PRC2 to test for colocalization of the rixosome and Polycomb bodies. We first validated each of the commercially available antibodies used in these experiments by showing that they recognized protein species that were depleted by specific siRNA treatments (Extended Data Fig. 3a–c). Consistent with the ChIP-seq results, immunofluorescence showed that MDN1 and WDR18 localized to closely overlapping domains with EZH2 Polycomb bodies (Extended Data Fig. 3d,e). The mammalian rixosome has previously been shown to localize to nucleoli, where it performs its rRNA processing functions<sup>41</sup>. To examine the relationship between Polycomb bodies and nucleoli, we stained cells for EZH2 and the nucleolar protein NPM1 and found that whereas the most intensely staining EZH2 foci co-localized with NPM1-stained nucleoli, the remaining EZH2 foci did not co-localize with NPM1 (Extended Data Fig. 3f,g).

## RING1A and B in rixosome recruitment

We next tested whether the localization of the rixosome to Polycomb target genes was Polycomb-dependent. As shown by heatmap analysis in Fig. 3a, the localization of both TEX10 and MDN1 to target loci was abolished in RING1A and RING1B (RING1A/B)-double-knockout (DKO) cells, whereas the levels of TEX10 and MDN1 were unaffected (Fig. 3b). Similarly, at the single-gene level, ChIP-seq signals for TEX10 and MDN1 on *PCDH10*, *IGFBP3* and HOXA genes were absent in the RING1A/B-DKO cells (Fig. 3c). Consistently, at the cytological level, the numbers of MDN1 and WDR18 foci were significantly reduced in RING1A/B-DKO or EZH1 and EZH2 (EZH1/2)-DKO cells, whereas the number of nucleoli (stained with NPM1) were unaffected (Extended Data Fig. 4a–c). Therefore, the localization of rixosome

subunits to both Polycomb target genes and Polycomb bodies required the catalytic PRC subunits.

**Fig. 3: PRC1 is required for rixosome chromatin targeting.**

 figure 3



**a**, Heatmap representations of ChIP-seq of TEX10 and MDN1 in HEK 293FT cells with the indicated treatments. Rank order is from highest to lowest TEX10 signal in control siRNA (siCtrl)-treated cells. RING1A/B-DKO cells were treated with control siRNA.  $\log_2$  enrichment levels were normalized to reads per genome coverage. Read counts per gene were summed in 50-nt bins. **b**, Immunoblots showing indicated protein levels in wild-type (WT) and RING1A/B-DKO cells. **c**, Genomic snapshots of ChIP-seq reads at Polycomb target genes *PCDH10*, *IGFBP3* and HOXA cluster for the indicated cells.  $\log_2$  enrichment levels were normalized to reads per genome coverage. **d**, Schematic of RING1B protein and its domains. CC, coiled-coil domain. The location of RING1B-2A substitutions is indicated. **e**, Immunoprecipitations showing the effect of RING1B-2A substitutions on the interaction of RING1B with PRC1 subunits CBX2, PHC2, BMI1 (PCGF4), RYBP, YAF2 and PCGF6, and rixosome subunits PELP1, TEX10 and SENP3 in HEK 293FT cells. **f**, ChIP-qPCR experiments showing the localization of MDN1 at the indicated genes in wild type, RING1B-2A, RING1B-KO and RING1A/B-DKO cell lines. Primers used for quantitative PCR targeted the first exon of each gene. *GAPDH* served as a control. Data are presented as mean  $\pm$  s.e.m. for three biological replicates. **g**, Immunoblots showing total H2AK119ub1 levels in wild-type, RING1B-2A and RING1A/B-DKO HEK 293FT cells. **h**, Heatmap representations of ChIP-seq of RING1B and H2AK119ub1 in wild-type, RING1B-2A HEK 293FT cells. Rank order is from highest to lowest RING1B signal in wild-type cells.  $\log_2$  enrichment levels were normalized to reads per genome coverage. Read counts per gene were summed in 50-nt bins.

#### [Source data](#)

To test whether the interaction of RING1B with the TEX10 subunit of the rixosome—observed with purified proteins (Extended Data Fig. 1*g–i*)—was required for the association of the rixosome with PRC1 and Polycomb target genes in cells, we used CRISPR–Cas9 to replace the chromosomal copies of RING1B with RING1B(Q137A/Q138A) (RING1B-2A) (Fig. 3*d*), which is impaired in its ability to bind to TEX10. As shown in Fig. 3*e*, the CBX2, BMI1 (also known as PCGF4) and PHC2 subunits of cPRC1, and the RYBP, YAF2 and PCGF6 subunits of vPRC1, co-immunoprecipitated with both wild-type and RING1B-2A proteins, indicating that the RING1B mutations did not disrupt the integrity of PRC1 complexes. However, whereas the TEX10, SENP3 and PELP1 subunits of the rixosome co-immunoprecipitated with wild-type RING1B, their interaction with RING1B-2A was greatly diminished (Fig. 3*e*). Consistent with the immunoprecipitation results, rixosome subunits NOL9, TEX10 and WDR18 co-migrated with PRC1 subunits PHC2 and RING1B during sucrose gradient sedimentation; this co-migration did not occur in extracts prepared from RING1B-2A mutant cells (Extended Data Fig. 5*a*). Moreover, experiments using chromatin immunoprecipitation followed by quantitative

PCR (ChIP–qPCR) showed that, relative to wild-type RING1B, the interaction of MDN1 with several Polycomb target loci was diminished to a similar extent in RING1B-2A and RING1B-knockout (KO) cells, whereas as expected, RING1A/B-DKO cells displayed a greater loss of MDN1 binding (Fig. 3f). Together, these results indicate that recruitment of the rixosome to target loci requires its specific interaction with RING1B. Consistent with maintenance of PRC1 integrity, RING1B-2A mutant cells had similar total levels of H2AK119ub1 to the wild type and ChIP–seq experiments showed that the genome-wide localization of RING1B itself and H2AK119 ubiquitination were not affected by RING1B-2A (Fig. 3g,h). Similarly, the depletion of NOL9 did not affect histone H2AK119ub1 or H3K27me3 levels, which were greatly diminished upon the depletion of RING1B and EZH2, respectively (Extended Data Fig. 5b–f). The rixosome therefore acts downstream of Polycomb-catalysed histone modifications.

## Rixosome regulates nascent RNA synthesis

We next investigated whether the rixosome was required for silencing of Polycomb target genes. We were unable to generate viable knockouts of several rixosome subunits, presumably owing to their essential roles in rRNA processing. We therefore used transient siRNA knockdown of rixosome subunits at timepoints that do not affect growth and proliferation to study the role of the rixosome in regulation of transcription. Growth curves after knockdown of the rixosome subunits NOL9 and LAS1L showed that 48 h of siRNA treatment did not affect cell proliferation (Extended Data Fig. 6a). We analysed changes in PolII levels and position at target genes by performing precision run-on sequencing (PRO-seq) 48 h after siRNA treatment. PRO-seq provides snapshots of transcriptionally engaged PolII with base-pair resolution<sup>42</sup>. In this way, we could focus on the direct transcriptional targets of the rixosome and Polycomb complexes, without the confounding effects of RNA processing or stability.

PRO-seq analysis revealed a significant increase in the PRO-seq signal of 228 genes and decreases in the PRO-seq signal of 30 genes in siNOL9 cells (adjusted *P* value ( $P_{adj}$ ) < 0.05; fold change > 1.5), and metagene analyses showed change in RNA polymerase signal at both TSSs and gene bodies (Fig. 4a,b, Extended Data Fig. 6b). To assess how this set of NOL9 target genes was affected by loss of RING1A/B or EED, we compared them to sets of expression-matched genes that were not affected by NOL9 depletion (Fig. 4b, Extended Data Fig. 6c). In contrast to siNOL9-unaffected or downregulated genes, siNOL9-upregulated genes were also mostly upregulated in siRING1A/B, RING1A/B-DKO and EED-KO cells (Fig. 4c, Extended Data Fig. 6d). Furthermore, relative to siNOL9-unaffected or -downregulated genes, siNOL9-upregulated genes showed highly significant enrichment in ChIP–seq signals for rixosome subunits (TEX10 and MDN1), a PRC1 subunit (RING1B), H2AK119ub1

and H3K27me3, but not H3K9me3 (Fig. 4d, Extended Data Fig. 6e–k). Consistently, relative to siNOL9-downregulated genes, we observed a greater overlap between siNOL9-upregulated genes and those also upregulated in RING1A/B-DKO and EED-KO cells (Extended Data Fig. 6l,m). As examples at the single-gene level, we observed increased PRO-seq signal at the *PCDH10*, *IGFBP3* and *HOXB6* genes in siNOL9 and RING1A/B-DKO cells (Fig. 4e). The increase in PolII occupancy in siNOL9 cells was in general weaker than in RING1A/B-DKO cells, which may be owing to partial depletion of NOL9 by siRNA treatment or additional rixosome-independent functions of RING1A/B.

**Fig. 4: Loss of rixosome upregulates Polycomb target genes at the level of nascent RNA synthesis.**

---

 **figure 4**

**a**, Average distribution of PRO-seq signal is shown at genes upregulated by siNOL9 ( $n = 228$ ). Data are shown in 25-nt bins. **b**, Average distribution of PRO-seq signal is shown at expression-matched genes unaffected by siNOL9 ( $n = 230$ ). Data are shown in 25-nt bins. **c**, Violin plots depict the  $\log_2$  fold change in PRO-seq for siNOL9

upregulated ( $n = 228$ ) and unaffected ( $n = 30$ ) genes in cells treated with siNOL9, siRING1B and RING1AB-DKO and EED-KO cells. Knockout cells were treated with control siRNA.  $P$  values are from two-tailed Mann–Whitney tests.  $P = 1.45 \times 10^{-5}$  for siRING1AB,  $P = 1.98 \times 10^{-17}$  for *RING1AB*-DKO and  $P = 3.39 \times 10^{-24}$  for EED-KO. **d**, Violin plots showing read counts for the indicated ChIP-seq experiments. Reads were summed  $\pm 1$  kb from the TSS for the gene groups indicated. Violin plots depict the range of values, with the centre line indicating the median.  $P$  values are from two-tailed Mann–Whitney tests.  $P = 1.55 \times 10^{-7}$  for TEX10,  $P = 0.0178$  for MDN1,  $P = 3.74 \times 10^{-10}$  for RING1B,  $P = 2.34 \times 10^{-8}$  for H2AK119ub1,  $P = 1.76 \times 10^{-11}$  for H3K27me3 and  $P = 0.0321$  for H3K9me3. **e**, Genome snapshots of PRO-seq experiments showing transcribing PolII at the indicated genes in siCtrl, siNOL9 and RING1A/B-DKO HEK 293FT cells. **f**, Venn diagrams showing overlap among upregulated (left) and downregulated (right) genes in cells treated with siNOL9 with upregulated genes in EED-KO and RING1A/B-DKO cells in RNA-seq experiments. Hypergeometric probability  $P$  values: siNOL9 upregulated versus RING1A/B-DKO,  $3.1 \times 10^{-122}$ ; siNOL9 upregulated versus EED-KO,  $3.6 \times 10^{-125}$ ; siNOL9 downregulated versus RING1A/B-DKO, 0.1; siNOL9 downregulated versus EED-KO,  $1.2 \times 10^{-3}$ ; RING1A/B-DKO versus EED-KO,  $4.2 \times 10^{-854}$ . **g**, Average distribution of normalized  $\log_2$  counts of the indicated ChIP-seq reads for genes that are upregulated (top) or downregulated (bottom) in HEK 293FT cells treated with siNOL9. Enrichment levels were normalized with reads per genome coverage. Read counts per gene were summed in 50-nt bins. **h**, Venn diagram showing overlap among upregulated genes in RING1B-2A, siLAS1L and RING1A/B-DKO cells in RNA-seq experiments; 1,143 genes (69%) were upregulated in both RING1B-2A-expressing and RING1A/B-DKO cells; 437 genes (79%) were overexpressed in both siLAS1L-treated and RING1A/B-DKO cells. Hypergeometric probability  $P$  values: RING1B-2A versus RING1A/B-DKO,  $1.2 \times 10^{-805}$ ; siLAS1L upregulated versus RING1A/B-DKO,  $2.2 \times 10^{-239}$ ; siLAS1L upregulated versus RING1B-2A,  $8.2 \times 10^{-157}$ . **i**, Dot plots showing changes in gene expression detected by RNA-seq of RING1B-2A cells, siLAS1L-treated cells and EED-KO cells in the sets of genes that are upregulated or downregulated in RING1A/B-DKO HEK 293FT cells. Data are mean  $\pm$  s.e.m.  $P$  value is from the two-tailed Mann–Whitney test.

## Source data

In agreement with the PRO-seq results, RNA-sequencing (RNA-seq) experiments showed that in contrast to siNOL9-downregulated genes, siNOL9-upregulated genes largely overlapped with genes upregulated in RING1A/B-DKO and EED-KO cells (Fig. 4f). Metagene analysis of genes affected in RNA-seq indicated that the siNOL9- and siLAS1L-upregulated, but not the downregulated genes were enriched for rixosome subunits and Polycomb-catalysed histone modifications (Fig. 4g, Extended Data Fig. 7a, b). Moreover, similar to RING1A/B-DKO, Polycomb target genes were

upregulated in cells expressing RING1B-2A (Fig. 4h). Additionally, as was the case with RING1A/B-DKO-upregulated genes, RING1B-2A-upregulated genes were enriched for rixosome subunits and Polycomb-catalysed histone modifications (Extended Data Fig. 7c). Together with the observations that RING1B-2A mutation or rixosome-subunit depletions did not affect H2AK119ub1 or H3K27me3 levels (Fig. 3g,h, Extended Data Fig. 5a–e), these results suggest that the rixosome and Polycomb complexes regulate a common set of genes at the level of transcription and that the rixosome acts downstream of Polycomb-catalysed histone modifications.

In agreement with the above analysis, RNA-seq experiments showed that the genes with increased PRO-seq signal in siNOL9, RING1A/B-DKO and EED-KO cells also had increased steady state RNA expression (Extended Data Fig. 7d). The set of siNOL9-upregulated genes in HEK 293FT cells included most of the HOX genes, which were also upregulated in RING1A/B-DKO but not in EED-KO cells (Extended Data Fig. 7e). This observation is consistent with the presence of H2AK119ub1 but little or no H3K27me3 at HOX genes in these cells (Fig. 2f). For example, genomic browser snapshots of RNA-seq reads showed that the depletion of either NOL9 or RING1A/B resulted in increased expression of the *PCHD10*, *IGFBP3* and *HOXB6* genes (Extended Data Fig. 7f). Notably, in contrast to wild-type cells, depletion of the rixosome subunits in EZH1/2-DKO or RING1A/B-DKO cells had no effect on the expression the *PCHD10* and several other target genes, indicating that the rixosome and Polycomb act epistatically through the same pathway (Extended Data Fig. 7g–i). As controls for possible indirect effects due to perturbation of ribosome biogenesis in the above experiments, siRNA knockdown of nucleolar NPM1 and PES1 proteins had no effect on the expression of several rixosome target genes (Extended Data Fig. 7g, i). Together, these results demonstrate that the rixosome and Polycomb complexes repress a largely shared set of genes in HEK 293FT cells.

## Rixosome functions in other cell types

To investigate the rixosome–Polycomb connection in other cell types, we examined the genome-wide localization of TEX10 in human embryonic stem (ES) cells and MDN1 in HeLa cells. Consistent with the results in HEK 293 cells, correlation and heatmap analysis of ChIP–seq reads indicated similar enrichment patterns for TEX10 (Extended Data Fig. 8a,b) and MDN1 (Extended Data Fig. 8c,d) with H2AK119ub1 and/or H3K27me3, but not with H3K9me, in human ES cells and HeLa cells. For example, TEX10 and MDN1, along with H2AK119ub1 and H3K27me3, co-localized to the entire HOXA cluster in human ES cells and HeLa cells (Extended Data Fig. 8e). Of note, 82% of TEX10-enriched TSSs in human ES cells and 76% of MDN1-enriched TSSs in HeLa cells overlapped with H2AK119ub1 peaks (Extended Data Fig. 8f,g). We also performed RNA-seq experiments in HeLa cells with siRNA knockdowns. As expected, we observed a high degree of correlation between the

genes that were upregulated upon the knockdown of NOL9, LAS1L and TEX10 rixosome subunits (Extended Data Fig. 9a). Moreover, consistent with the results in HEK 293 cells, a large fraction of the genes upregulated in siNOL9, siLAS1L, and siTEX10 cells overlapped with those upregulated in RING1A-KO, siRING1B or siEZH2 cells (Extended Data Fig. 9b–e). As in HEK 293 cells, depletion of rixosome subunits in HeLa cells resulted in increased expression of HOX genes, which were also upregulated in siEZH2 cells, consistent with their association with H3K27me3 in these cells (Extended Data Fig. 8e, Extended Data Fig. 9f). Similarly, metagene analysis indicated that the upregulated, but not downregulated genes were enriched for rixosome subunits and Polycomb-catalysed histone modification (Extended Data Fig. 9g,h). The larger overlap between rixosome- and Polycomb-repressed genes in HeLa and human ES cells is probably owing to differences in siRNA knockdown efficiencies in these cells and/or in regulatory strategies. The rixosome therefore contributes to Polycomb silencing in different cell types.

## Rixosomal RNA degradation and silencing

The rixosome contains RNA endonuclease and polynucleotide kinase activities that prepare target RNAs for further degradation by the 5'-3' XRN2 exoribonuclease<sup>6,43</sup> (Fig. 5a). Cleavage of target RNA by the LAS1L endoribonuclease subunit of the rixosome generates a 5'-OH group, which must be phosphorylated by the NOL9 polynucleotide kinase subunit for the RNA to become a substrate for degradation by XRN2<sup>6</sup> (Fig. 5a). We performed depletion and rescue experiments to first test whether the enzymatic activities of each LAS1L and NOL9 were required for their silencing functions. The upregulation of several target genes by the depletion of either LAS1L or NOL9 was rescued by the reintroduction of siRNA-resistant wild-type (WT) versions (Flag–LAS1L(WT) or haemagglutinin (HA)–NOL9(WT)) but not their catalytically dead mutant versions<sup>44,45</sup> (Flag–LAS1L-2A or HA–NOL9(K312A)) (Fig. 5b,c). The requirements for the endonuclease activity of LAS1L and the polynucleotide kinase activity of NOL9, respectively, strongly suggest that the rixosome mediates target RNA degradation via the XRN2 exoribonuclease (Fig. 5a). To test this hypothesis, we knocked down XRN2 with two different siRNAs and found that several targets of the rixosome and Polycomb pathways were expressed at elevated levels in the knockdown cells, whereas three non-target loci were not affected (Fig. 5d; see Extended Data Fig. 10a–c for knockdown validation). Furthermore, the silencing defects resulting from XRN2 depletion were rescued by wild-type (Flag–XRN2(WT)) but not a catalytically dead<sup>46</sup> XRN2 (Flag–XRN2(E203G)). We therefore conclude that the rixosome and XRN2 work together to degrade RNA at Polycomb target loci.

**Fig. 5: Rixosome-associated RNA degradation is required for repression of Polycomb-regulated genes.**

---

 **figure 5**

**a**, Schematic of rixosome RNA-processing activities depicting the hypothesis that the LAS1L endonuclease and RNA NOL9 kinase activities are required to prepare target RNA for XRN2-mediated degradation. **b–d**, Quantitative PCR with reverse transcription (RT–qPCR) analysis of expression of the indicated Polycomb target

---

genes in the indicated siRNA-treated HEK 293FT cells rescued with LAS1L (**b**), NOL9 (**c**) or XRN2 (**d**). Expression levels were normalized to *ACTB* and siCtrl-treated cells. Data are presented as mean ± s.e.m. for three biological replicates. **e**, Schematic showing construction of cell lines with a 5 $\times$ tetO-H2B-CITRINE (H2B-CTRN) reporter gene expressing rTetR–RING1B wild type or rTetR–RING1B-2A fusion proteins. **f–h**, ChIP–qPCR analysis of H2B-CTRN reporter enrichment for RING1B (**f**), H2AK119ub1 (**g**) and TEX10 (**h**) following 21 days of doxycycline treatment. ChIP–qPCR signals were normalized to *GAPDH*. Dots represent individual biological replicates. **i**, RT–qPCR analysis showing H2B-CTRN reporter RNA expression in rTetR–RING1B-expressing wild-type or RING1B-2A HEK 293FT cells before treatment relative to after 21 days of doxycycline treatment. RNA expression levels were normalized to *ACTB*. Data are presented as mean ± s.e.m. for three biological replicates. **j**, RT–qPCR analysis of H2B-CTRN RNA expression in the indicated siRNA-treated and NOL9-rescued HEK 293FT cells 3 days after release from 21-day doxycycline treatment. Expression levels were normalized to *ACTB* and to siCtrl-treated cells. Data are presented as mean ± s.e.m. for three biological replicates **k**, Model for the role of rixosome in Polycomb-mediated gene silencing. The rixosome is recruited through the interaction of RING1B in the vPRC1 and PRC1 complexes with the TEX10 subunit of the rixosome (individual subunits not shown) to mediate nascent RNA degradation and transcription termination. The rixosome also interacts with PRC2. See text for details.

#### [Source data](#)

## Ectopic RING1B can recruit the rixosome

To provide further evidence that RING1B could recruit the rixosome to chromatin, we fused wild-type or RING1B-2A to the bacterial reverse tetracycline repressor (resulting in rTetR–RING1B or rTetR–RING1B-2A) and tested whether they could recruit TEX10 to TetR binding sites inserted together with a reporter gene at a euchromatic locus (5 $\times$ tetO-CTRN) (Fig. 5e). ChIP–qPCR experiments showed that both wild-type RING1B and RING1B-2A mutant proteins were recruited to the ectopic locus and induced similar levels of H2AK119ub1 (Fig. 5f,g). However, whereas wild-type RING1B recruited high levels of TEX10 to the ectopic locus, RING1B-2A, which does not interact with TEX10 or the rixosome, recruited little or no TEX10 (Fig. 5h). The low levels of TEX10 recruitment induced by RING1B-2A are probably mediated by binding of the endogenous wild-type RING1B (PRC1) to H2AK119ub1 at the ectopic locus. Therefore, consistent with biochemical data and in vivo analysis of the requirements for rixosome localization to Polycomb target genes, these results demonstrate that RING1B can directly recruit the rixosome to chromatin.

We next tested the effect of RING1B tethering and depletion of NOL9 on the expression of the *5xtetO-CTRN* reporter. In the presence of doxycycline, which induces strong binding of rTetR fusion proteins to *5xtetO* sites, we observed several-hundred-fold repression of *CTRN* reporter RNA for both rTetR–RING1B and rTetR–RING1B-2A tethering (Fig. 5*i*). However, depletion of NOL9 resulted in only weak derepression of the reporter (Extended Data Fig. 10*d*). We reasoned that the continuous strong binding of rTetR–RING1B to the reporter locus may partially mask the requirement for the rixosome. To test this hypothesis, we performed siRNA depletion experiments three days after the release of rTetR–RING1B (removing doxycycline from the medium). Under these conditions, depletion of NOL9 resulted in strong derepression of the *CTRN* reporter, which was rescued by wild-type but not catalytically dead NOL9 (Fig. 5*j*). As controls, depletion of RING1B, but not NPM1, resulted in strong derepression of the reporter gene in both the absence and presence of doxycycline (Fig. 5*j*, Extended Data Fig. 10*d*). Therefore, similar to endogenous loci, RING1B-mediated rixosome recruitment contributes to silencing at the ectopic locus.

## Discussion

Our findings demonstrate a role for the conserved rRNA processing and ribosome biogenesis complex, the rixosome, in Polycomb-mediated gene silencing. We demonstrate that the rixosome is recruited to chromatin in a PRC1-dependent manner by binding to RING1B and our PRO-seq analysis of active transcription shows that many genes targeted by these pathways contain paused PolII downstream of their promoter regions. Upon the loss of either the rixosome or Polycomb, the density of both the paused and elongating polymerase at these target genes increases, suggesting that Polycomb-mediated rixosome recruitment blocks productive transcription elongation by paused and/or elongating polymerase, thereby repressing gene activity.

In one model, silencing by Polycomb complexes is thought to involve chromatin compaction to block transcription initiation<sup>1,2</sup>. In both flies and mammals, subunits of the PRC1 complex can condense nucleosomal arrays in vitro and in vivo<sup>7,22,23,24</sup> and, in mammals, PRC2 alone has in vitro chromatin compaction activity<sup>25,26</sup>. Moreover, recent studies show that the CBX2 subunit of the cPRC1 complex, which mediates its chromatin compaction activity, also promotes liquid–liquid phase separation in vitro and in vivo<sup>27,28,29</sup>. Our identification of a role for the rixosome in silencing of Polycomb target genes suggests that an additional layer of regulation involving RNA degradation has an important role in silencing of Polycomb target genes (Fig. 5*k*). We propose that the rixosome, once recruited to repressed genes by PRC1 and/or PRC2 complexes, surveys these loci for the presence of nascent RNA. At loci where Polycomb-mediated repression is weak and PolII enters early elongation, the rixosome recognizes and associates with nascent RNA to process it for degradation (Fig. 5*k*). Accordingly, we provide evidence that rixosome-cleaved RNAs become substrates for

the 5'-3' exoribonuclease XRN2, suggesting a role for nascent RNA cleavage and transcription termination in the potent silencing of Polycomb target genes. Heterochromatin-associated RNA degradation appears to have diverse and broadly conserved roles in gene silencing. In fission yeast, plants and animals, RNA interference-dependent and -independent RNA degradation contributes to heterochromatin establishment and maintenance<sup>5,47</sup>. More recently, the LSM2-8 RNA decapping complex has been reported to act together with the XRN2 exonuclease to ensure full silencing of H3K27me3 loci in *Caenorhabditis elegans*<sup>48</sup>, suggesting that distinct mechanisms may act upstream of XRN2-mediated RNA degradation at Polycomb loci.

The rixosome may also regulate how chromatin-associated RNAs affect other Polycomb functions. PRC2 has been shown to interact with RNA promiscuously and with high affinity<sup>49,50</sup>, and RNA has been suggested to have both positive and negative roles in promoting the association of PRC2 with chromatin<sup>51,52,53</sup>. Rixosome-mediated RNA degradation may coordinate the different effects of RNA, particularly if the positive and negative roles of RNA were temporally regulated. The roles of the rixosome in the silencing functions of different types of chromatin, constitutive H3K9me heterochromatin in fission yeast<sup>4</sup> and facultative H2AK119ub1 and H3K27me3 heterochromatin in human cells (this study), suggest that its RNA-degradation activities have highly conserved and critical functions in heterochromatin-mediated gene silencing.

## Methods

### Plasmid construction

Rixosome subunits (NOL9, WDR18, PELP1, TEX10), PRC2 subunits (EZH2, EED, SUZ12), PRC1 subunits (RING1B, PCGF1-4) and CBX1-8 cDNAs were amplified from human ES cell cDNA library and inserted to pGAD-T7 (Takara, 630442) and pGBK-T7 (Takara, 630443) plasmids for Y2H assays. NOL9 siRNA resistant cDNA was generated by PCR. The siRNA target sequence was mutated from 5'-AGACCTAACGTTCTGCGAA-3' to 5'-CGGCCGAAATTTGCAGGA-3' and integrated into the pCI (Promega, E1731) plasmid for ectopic protein expression. For bacteria protein expression, cDNA was integrated to pGEX-6P-1 (GE Healthcare, 28-9546-48).

### Y2H assays

Y2H budding yeast strain (Takara) was cultured with YEPD+adenine overnight at 30 °C. Yeast cells were collected OD 0.5 by centrifugation at 3,000 rpm for 3 min. Cells

were resuspended and washed 2 times with 0.1 M LiAc (in 1x TE buffer). The bait pGBK7 (0.5 µg) expressing rixosome, Polycomb, and HP1proteins and prey pGADT7 (0.5 µg) vectors were mixed with 10 µg carrier DNA, and further mixed with yeast cells collected from 10-ml cultures and resuspended in 50 µl 0.1 M LiAc (in 1× TE buffer). DNA-yeast mixture was incubated with 130 µl 40% PEG 4000 for 30 min at 30 °C. For transformation, 21 µl DMSO was added and mixed well with the yeast–DNA mixture, followed by heat shock at 42 °C for 20 min. After incubation on ice for 3 min, the cells were pelleted by centrifugation for 3 min at 4 °C. The supernatant was then discarded and sterile water was added to resuspend the cells, which were plated on double selective medium SC plates (Trp-, Leu-) for 3 days at 30 °C. Colonies were further transferred to quadruple selective medium SC plates (Trp-, Leu-, His-, Ade-) for 3–4 days at 30 °C. For spotting assays, cells were incubated overnight in 4 ml double selective SC medium (Trp-, Leu-). The cells were then diluted to an optical density at 600 nm of 1, one millilitre of which was pelleted, washed once with sterilized water, resuspended in 250 µl sterilized water, and transferred to 96-well plates. Three microlitres of cell suspension from each well was plated on double-selective medium SC plates (Trp-, Leu-) and quadruple-selective medium SC plates (Trp-, Leu-, His-, Ade-) for four days.

## Cell culture

HeLa (ATCC, CCL-2), and HEK 293FT (ThermoFisher, R70007) cells were cultured in DMEM containing 10% fetal calf serum. Human embryonic stem cells were authenticated by the Initiative for Genome Editing and Neurodegeneration of Harvard Medical School and cultured as previously described<sup>54</sup>. In brief, cells cultures on 0.08 mg ml<sup>-1</sup> matrigel coated plates with DMEM/F12 (containing 5 µg ml<sup>-1</sup> insulin and 10 µg ml<sup>-1</sup>, 0.1 µg ml<sup>-1</sup> FGF2, 1.7 ng ml<sup>-1</sup> TGFβ1, 10 µg ml<sup>-1</sup> transferrin). Cells were tested for mycoplasma contamination by the suppliers and were negative.

## RNAi

For siRNA-mediated knockdown, Lipofectamine RNAiMAX transfection reagent (Invitrogen) and siRNA (200 nM) were used to transfect the cells by following the manufacturer’s instructions. All the siRNAs were synthesized by Dharmacon and are listed in Supplementary Table 1.

## CRISPR–Cas9-mediated human genome editing

Small guide RNA was synthesized via in vitro transcription by using MAXIscript T7 transcription kit (ThermoFisher, AM1312). CRISPR–Cas9 protein was purified by the Initiative for Genome Editing and Neurodegeneration Core in the Department of Cell Biology at Harvard Medical School. DNA Oligonucleotide templates (synthesized by

IDT, Supplementary Table 2), guide RNA, and CRISPR–Cas9 protein were delivered to cells by electroporation with Neon transfection system (ThermoFisher). Clones were screened by PCR and Miseq sequencing (Illumina).

## Immunofluorescence

Cells were placed on plates with cover slides. Cells were first washed with PBS, and fixed and permeabilized with methanol for 8 min at  $-20^{\circ}\text{C}$ . Cells were then incubated for 4–10 h at  $4^{\circ}\text{C}$  with primary antibodies in PBS containing 4% bovine serum, which was followed by staining with secondary antibodies and  $1 \mu\text{g ml}^{-1}$  DAPI. A confocal microscope (Nikon, Ti with perfect focus and spinning disk) equipped with a  $60\times/1.40$  NA objective lens was used to image cells. NIS-Elements imaging software was used for imaging data collection. Images were post-processed with ImageJ (NIH) and photoshop (Adobe) software. EZH2 and MDN1 fluorescence intensities were assessed using ImageJ. NPM1 foci were counted visually directly using ImageJ. For MDN1 foci, the signal was measured in the regions with NPM1 in control cells, foci with the lowest value of NPM1 staining in the control cells was then used as a cutoff and any foci measured by ImageJ with higher value were counted as MDN1 foci. A list of antibodies and their sources is described in Supplementary Table 3.

## Immunoprecipitation and mass spectrometry analysis

To prepare chromatin-enriched fractions, cells were washed with PBS and then resuspended in ice-cold hypotonic buffer (10 mM HEPES, pH7.9, 1.5 mM MgCl<sub>2</sub>, 10 mM KCl, 0.2 mM PMSF, 0.2 mM DTT) and incubated on ice for 10 min. Cell membranes were then disrupted by douncing 10 times. Nuclei were pelleted by centrifugation at 2,000g for 10 min, resuspended in cell lysis buffer (50 mM Hepes, pH 7.4, 150 mM NaCl, 1 mM MgCl<sub>2</sub>, 1 mM EGTA, and 0.5% Triton X-100) by pipetting for 3 min, and pelleted by centrifugation at 2,000g for 10 min to obtain a chromatin fraction. The chromatin pellet was resuspended in IP buffer (50 mM Hepes, pH 7.4, 250 mM NaCl, 10% glycerol, 1 mM MgCl<sub>2</sub>, 1 mM EGTA, and 1% Triton X-100) containing protease inhibitor cocktail (5056489001, Sigma) and 1 mM DNase I. Chromatin was digested for 2 h at  $4^{\circ}\text{C}$  and centrifuged at 10,000g for 10 min. The supernatant was then incubated with specific antibodies (Supplementary Table 3) and immune complexes were collected using Dynabeads Protein A/G (ThermoFisher). For silver staining, samples were run on a 5%–20% Bis-Tris SDS–PAGE gel (BioRad) and stained with SilverQuest Silver Staining kit (Invitrogen) according to the manufacturer's instructions. For immunoblotting, beads were boiled for 5 min in SDS loading buffer. For immunoprecipitations in Fig. 1f,g, Benzonase (Sigma, E8263) treatment was performed by adding  $500 \text{ U ml}^{-1}$  benzonase to cell lysates followed by incubation for 1 h in  $4^{\circ}\text{C}$  before incubation with antibody immobilized beads. For

mass spectrometry analysis, proteins were eluted with 0.5 M NH<sub>4</sub>OH and dried to completion in a speed vac.

For Flag–NOL9 and Flag–WDR18 immunoprecipitation and mass spectrometry, dried protein samples were digested in 200 mM EPPS buffer pH 8.5 with trypsin (Promega V5111). Digests contained 2% acetonitrile (v/v) and were performed at 37 °C overnight. Digests were labelled directly with TMT10 plex reagents (ThermoFisher Scientific, 90406). Labelling efficiency was checked by mass spectrometry. After hydroxylamine-quenching (0.3% v/v) for 15 min, reactions were mixed and acidified and solvent evaporated to near completion by speed vac. Samples were then fractionated by alkaline reversed phase chromatography (ThermoFisher 84868) into 12 fractions eluted with 10%, 12.5%, 15%, 17.5%, 20%, 25%, 30%, 35%, 40%, 50%, 65% and 80% acetonitrile. Fractions were pooled into 6 fractions (1+7, 2+8, 3+9, 4+10, 5+11, 6+12), dried down, stage-tipped and analysed by mass spectrometry on an Orbitrap Lumos instrument (Thermo Scientific). Relative quantification followed a multi-notch SPS-MS<sup>3</sup> method. Prior to injection, peptides were separated by HPLC with an Easy-nLC 1200 liquid chromatography system using 100 μm inner diameter capillaries and a C<sub>18</sub> matrix (2.6 μM Accucore C<sub>18</sub> matrix, ThermoFisher Scientific). Peptides were separated with 4-hour acidic acetonitrile gradients. MS<sup>1</sup> scans were measured by orbitrap recording (resolution 120,000, mass range 400–1400 Th). After collision induced dissociation (CID) (35%), MS<sup>2</sup> spectra were collected by iontrap mass analyser. After SPS (synchronous precursor selection), TMT reporter ions were generated by high-energy collision-induced dissociation (HCD) (55%) and quantified by orbitrap MS<sup>3</sup> scan (resolution 50,000 at 200 Th). Spectra were searched with an in-house written software based on Sequest (v.28, rev. 12) against a forward and reversed human proteome database (Uniprot 07/2014). Mass tolerance for searches was 50 ppm for precursors and 0.9 Da for fragment ions. Two missed tryptic cleavages per peptide were allowed and oxidized methionine (+15.9949 Da) was searched dynamically. For a peptide FDR (false discovery rate) of 1%, a decoy database strategy and linear discriminant analysis (LDA) were applied. The FDR for collapsed proteins was 1%. Proteins were quantified by summed peptide TMT s/n (signal/noise) with a sum s/n > 200 and an isolation specificity of >70%. Details of the TMT workflow and sample preparation procedures were described recently [55](#).

For Flag–PHC2 and Flag–CBX4 immunoprecipitation and mass spectrometry, we added 20 μl of 8 M urea, 100 mM EPPS pH 8.5 to the beads. We added 5 mM TCEP and incubated the mixture for 15 min at room temperature. We then added 10 mM of iodoacetamide for 15 min at room temperature in the dark. We added 15 mM DTT to consume any unreacted iodoacetamide. We added 180 μl of 100 mM EPPS pH 8.5. to reduce the urea concentration to <1 M, 1 μg of trypsin, and incubated at 37 °C for 6 h. The solution was acidified with 2% formic acid and the digested peptides were desalted via StageTip, dried via vacuum centrifugation, and reconstituted in 5%

acetonitrile, 5% formic acid for LC-MS/MS processing. All label-free mass spectrometry data were collected using a Q Exactive mass spectrometer (Thermo Fisher Scientific) coupled with a Famos Autosampler (LC Packings) and an Accela600 liquid chromatography (LC) pump (Thermo Fisher Scientific). Peptides were separated on a 100 µm inner diameter microcapillary column packed with about 20 cm of Accucore C18 resin (2.6 µm, 150 Å, Thermo Fisher Scientific). For each analysis, we loaded about 2 µg onto the column. Peptides were separated using a 1 h method from 5 to 29% acetonitrile in 0.125% formic acid with a flow rate of about 300 nl min<sup>-1</sup>. The scan sequence began with an Orbitrap MS1 spectrum with the following parameters: resolution 70,000, scan range 300–1,500 Th, automatic gain control (AGC) target  $1 \times 10^5$ , maximum injection time 250 ms, and centroid spectrum data type. We selected the top twenty precursors for MS2 analysis which consisted of HCD high-energy collision dissociation with the following parameters: resolution 17,500, AGC  $1 \times 10^5$ , maximum injection time 60 ms, isolation window 2 Th, normalized collision energy (NCE) 25, and centroid spectrum data type. The underfill ratio was set at 9%, which corresponds to a  $1.5 \times 10^5$  intensity threshold. In addition, unassigned and singly charged species were excluded from MS2 analysis and dynamic exclusion was set to automatic. Mass spectrometric data analysis. Mass spectra were processed using a Sequest-based in-house software pipeline. MS spectra were converted to mzXML using a modified version of ReAdW.exe. Database searching included all entries from the *S. pombe* UniProt database which was concatenated with a reverse database composed of all protein sequences in reversed order. Searches were performed using a 50 ppm precursor ion tolerance. Product ion tolerance was set to 0.03 Th. Carbamidomethylation of cysteine residues (+57.0215 Da) were set as static modifications, while oxidation of methionine residues (+15.9949 Da) was set as a variable modification. Peptide spectral matches (PSMs) were altered to a 1% FDR. PSM filtering was performed using a linear discriminant analysis, as described previously, while considering the following parameters: XCorr, ΔCn, missed cleavages, peptide length, charge state, and precursor mass accuracy. Peptide-spectral matches were identified, quantified, and collapsed to a 1% FDR and then further collapsed to a final protein-level FDR of 1%. Furthermore, protein assembly was guided by principles of parsimony to produce the smallest set of proteins necessary to account for all observed peptides.

## GST pulldown and immunoblotting

Proteins for GST pulldown assays were expressed in BL21 Codon Plus *Escherichia coli* (Agilent Technologies) with 200 µM IPTG induction at 16 °C overnight. Bacteria were then collected and washed with cold PBS, and sonicated (Branson sonicator) for 1 min with 20% amplitude at 4 °C. Sonicated samples were centrifuged at 20,000g for 10 min, and the supernatant was added to 0.5 ml Glutathione Sepharose 4B resin (GE Healthcare, 17075605), which was equilibrated with PBS. GST-tagged proteins were

incubated with the resin for 2 h at 4 °C. The resin was then washed 6 times with PBS containing 1% Triton 100. To remove the GST tag, bead-coupled proteins were incubated with PreScission Protease (GE Healthcare, 27-0843-01) in reaction buffer (50 mM Tris-HCl, pH7.0, 150 mM NaCl, 1 mM EDTA, 1 mM DTT) for 2 h at 4 °C. The GST-tagged PreScission Protease was removed using Glutathione Sepharose 4B resin.

For GST pulldown assays, 10 µl 50% slurry of Glutathione Sepharose 4B was used for each sample. GST or GST-tagged proteins (0.1 µM) were incubated with untagged proteins (0.1 µM) in 1 ml PBS (137 mM NaCl, 2.7 mM KCl, 8 mM Na<sub>2</sub>HPO<sub>4</sub>, and 2 mM KH<sub>2</sub>PO<sub>4</sub>, pH7.4) containing 0.5% Triton 100 overnight at 4 °C. Beads were washed 3 times with PBS containing 0.5% Triton 100, resuspended in SDS protein buffer, and boiled for 5 min. Input (2–5%) and bound proteins (10–50%) were run on 4–20% gradient SDS–PAGE gel. SDS–PAGE was performed to separate proteins for 2 h at 80 V, and proteins were transferred to a PVDF membrane (Millipore). The membranes were blocked in 3% milk in PBS with 0.2% Tween-20, and sequentially incubated with primary antibodies and HRP-conjugated secondary antibodies, or directly incubated with HRP-conjugated primary antibodies for chemiluminescence detection. Sources of antibodies can be found in Supplementary Table 3.

### Sucrose gradient centrifuge fractionation assay

Flag-tagged proteins were purified from the soluble chromatin fraction using magnetic beads (Sigma, M8823) and eluted with 3×Flag peptides (APExBIO, A6001) in elution buffer (20 mM Hepes-KOH, pH7.5, 100 mM KOAc, 5 mM Mg(OAc)<sub>2</sub>, 1 mM EDTA, 10% Glycerol). Sucrose gradients (10%-30%) were prepared using the Gradient station (BIOCOP). An Optima TLX Ultracentrifuge equipped with TLS-55 rotor was used for ultracentrifugation for 16 h at 4 °C with 35 k rpm. Gradients of 2.2 ml were fractionated into 22 fractions. One-hundred-microlitre fractions were pipetted from top and protein in fractions was captured using StrataClean resin (Agilent, 400714). Protein samples were boiled in SDS sample buffer (62.5 mM Tris-HCl, pH 6.8, 2% SDS, 10% glycerol, 0.01% bromophenol blue) for 3 min at 98 °C, and analysed by immunoblotting following gel electrophoresis (4%–15% precast protein gel with SDS from Biorad, 4561081).

### RT-qPCR

Total RNA was extracted using the RNeasy Plus kit (74134, Qiagen) and reverse transcribed into cDNA using gene-specific primers and reverse transcription kit (18090010, ThermoFisher). cDNA was analysed by running PCR on a QuantStudio 7 Flex Real Time PCR System (Applied Biosystem). All reactions were performed using 10 ng RNA in a final volume of 10 µl. PCR parameters were 95 °C for 2 min and 40

cycles of 95 °C for 15 s, 60 °C for 15 s, and 72 °C for 15 s, followed by 72 °C for 1 min. All the quantitative PCR data presented were at least three biological replicates. The forward and reverse primers used for RT–qPCR targeted the first exons of the genes. Primer sequences are presented in Supplementary Table 4.

## RNA-seq

Total RNA was isolated from human cells with an RNA purification kit (Qiagen, 74134) and genomic DNA was removed by genomic DNA binding columns in the kit. Two micrograms of total RNA was used for RNA-seq library construction. Poly(A)-containing mRNA was isolated by poly(A) selection beads and further reverse transcribed to cDNA. The resulting cDNA was ligated with adapters, amplified by PCR, and further cleaned to obtain the final library. Libraries were sequenced on an Illumina Hiseq machine (Novogene) to obtain 150 bp paired-ended reads.

RNA-seq reads were pseudo aligned using Kallisto 0.45.1. An index was generating using the Ensembl hg19 GTF and cDNA FASTA. Kallisto was run using default parameters with two exceptions: allowing searching for fusions ( $-f$ usion) and setting bootstrap to 100 ( $-b$  100).

To visualize the mapped RNA-seq with IGV or UCSC genome browser, bam files were generated with Hisat 2.2.0, which was followed by making bigwig files with deeptools (v3.0.2) (binsize 10). Reads were normalized to reads per genome coverage.

Read counts were calculated on a per transcript basis using Kallisto and the above described pseudoalignment. The R package tximport 1.10.1 was used to select the dominant transcript per gene (`txOut = FALSE`), which was then used for DEseq2 analysis. To analyse only active genes, those with 0 read counts in all samples were removed from the DEseq2 output. As they are not transcribed by PolII, 13 genes on chrM were also removed, resulting in a list of 24,043 active genes. Upregulated genes and downregulated genes are defined with  $P_{adj} < 0.05$  and fold change  $> 2$  or  $< -2$ .

## PRO-seq library construction

Aliquots of frozen ( $-80^{\circ}\text{C}$ ) permeabilized cells were thawed on ice and pipetted gently to fully resuspend. Aliquots were removed and permeabilized cells were counted using a Luna II, Logos Biosystems instrument. For each sample, 1 million permeabilized cells were used for nuclear run-on, with 50,000 permeabilized *Drosophila* S2 cells added to each sample for normalization. Nuclear run on assays and library preparation were performed essentially as described<sup>56</sup> with modifications noted: 2× nuclear run-on buffer consisted of (10 mM Tris (pH 8), 10 mM MgCl<sub>2</sub>, 1

mM DTT, 300 mM KCl, 40 µM each biotin-11-NTPs (Perkin Elmer), 0.8 U µl<sup>-1</sup> SuperaseIN (Thermo), 1% sarkosyl). Run-on reactions were performed at 37 °C. Adenylated 3' adapter was prepared using the 5' DNA adenylation kit (NEB) and ligated using T4 RNA ligase 2, truncated KQ (NEB, per manufacturer's instructions with 15% PEG-8000 final) and incubated at 16 °C overnight. One-hundred-eighty microlitres of betaine blocking buffer (1.42 g of betaine brought to 10 ml with binding buffer supplemented to 0.6 µM blocking oligonucleotide (TCCGACGATCCCACGTTCCCGTGG/3InvdT/)) was mixed with ligations and incubated 5 min at 65 °C and 2 min on ice prior to addition of streptavidin beads. After T4 polynucleotide kinase (NEB) treatment, beads were washed once each with high salt, low salt, and blocking oligonucleotide wash (0.25× T4 RNA ligase buffer (NEB), 0.3 uM blocking oligonucleotide) solutions and resuspended in 5' adapter mix (10 pmol 5' adapter, 30 pmol blocking oligonucleotide, water). 5' adapter ligation was per Reimer but with 15% PEG-8000 final. Eluted cDNA was amplified with five cycles (NEBNext Ultra II Q5 master mix (NEB) with Illumina TruSeq PCR primers RP-1 and RPI-X) following the manufacturer's suggested cycling protocol for library construction. A portion of preCR was serially diluted and for test amplification to determine optimal amplification of final libraries. Pooled libraries were sequenced using the Illumina NovaSeq platform.

## PRO-seq data analysis

All custom scripts described herein are available on the Adelman Lab Github ([https://github.com/AdelmanLab/NIH\\_scripts](https://github.com/AdelmanLab/NIH_scripts)). Using a custom script (trim\_and\_filter\_PE.pl), FASTQ read pairs were trimmed to 41 bp per mate, and read pairs with a minimum average base quality score of 20 retained. Read pairs were further trimmed using cutadapt 1.14 to remove adapter sequences and low-quality 3' bases (-match-read-wildcards -m 20 -q 10). R1 reads, corresponding to RNA 3' ends, were then aligned to the spiked in Drosophila genome index (dm3) using Bowtie 1.2.2 (-v 2 -p 6-best-un), with those reads not mapping to the spike genome serving as input to the primary genome alignment step (using Bowtie 1.2.2 options -v 2-best). Reads mapping to the hg19 reference genome were then sorted, via samtools 1.3.1 (-n), and subsequently converted to bedGraph format using a custom script (bowtie2stdBedGraph.pl). Because R1 in PRO-seq reveals the position of the RNA 3' end, the '+' and '-' strands were swapped to generate bedGraphs representing 3' end position at single nucleotide resolution.

For NOL9 KD PRO-seq, we performed 2 sets of PRO-seq experiments, each with two biological replicates. In the first set of experiments, NOL9 depletion resulted in many more upregulated (228) than downregulated (30) genes, while in the second set experiments, nearly the same number of genes were up (162) and down (160) regulated. Furthermore, unlike the first set, in the second set, the extent of overlap

between siNOL9 upregulated and downregulated genes with those upregulated in EED-KO or RING1A/B-DKO was similar. Although the basis of this discrepancy is unclear, the correlation between the two biological replicates in Set2 was lower than Set1 raising the possibility that poor growth or inefficient NOL9 depletion in Set2 siNOL9 cells may have resulted in a larger number of non-specifically downregulated genes. We therefore eliminated the Set2 siNOL9 data and used only the 2 biological replicates from the Set1 siNOL9 experiment.

## Gene model refinement using PRO-seq and RNA-seq

To select gene-level features for differential expression analysis, as well as for pairing with PRO-seq data, we assigned a single, dominant TSS and transcription end site (TES) to each active gene. This was accomplished using a custom script, get\_gene\_annotations.sh (available at

<https://github.com/AdelmanLab/GeneAnnotationScripts>), which uses RNA-seq read abundance and PRO-seq R2 reads (RNA 5' ends) to identify dominant TSSs, and RNA-seq profiles to define most commonly used TESs. RNA-seq and PRO-seq data from control and siNOL9 cells were used for this analysis, to capture gene activity under both conditions. Exon- and transcript-level features consistent with the resulting TSS to TES windows for 21,004 active genes in HEK 293T cells were selected from an hg19 reference GTF (GRCh38.99 from Ensembl). This filtered list of active genes was used for analyses shown in Figs. 2c–e, 4a–d, Extended Data Figs, 2c, d, 6b–k, as well as for defining differentially expressed genes in PRO-seq data. Differentially expressed genes between control ( $n = 2$ ) and siNOL9 ( $n = 2$ ) cells were determined using DESeq2 v1.26.0. Genes were called as differentially expressed using DEseq2's DESeqDataSetFromMatrix mode at an adjusted  $P$  value threshold of  $<0.05$  and fold change  $>1.5$ . This revealed 228 genes to be upregulated and 30 genes to be downregulated upon siNOL9.

## ChIP–qPCR, ChIP–seq and data analysis of ChIP–seq

ChIP was performed as previously described with minor modifications<sup>57</sup>. Cells for ChIP were cultured in 15 cm plates. Cell were first washed with cold PBS, crosslinked at room temperature with 10 mM DMP (ThermoFisher Scientific) for 30 min, and then 1% formaldehyde (ThermoFisher Scientific) for 15 min. Crosslinking reactions were quenched by addition of 125 mM glycine for 5 min. Crosslinked cells were separated by 3 min treatment of 0.05% trypsin (Gibco), and then washed with cold PBS 3 times. In every wash, cells were centrifuged for 3 min at 1,000g at 4 °C. Cell were then resuspended in sonication buffer (pH 7.9, 50 mM Hepes, 140 mM NaCl, 1 mM EDTA, 1% Triton, 0.1% Sodium deoxycholate, and 0.5% SDS) and sonicated to shear chromatin into ~300 bp fragments using a Branson sonicator. Sonicated samples were diluted fivefold with ChIP dilution buffer (pH 7.9, 50 mM Hepes, 140 mM NaCl, 1

mM EDTA, 1% Triton, 0.1% Sodium deoxycholate) to obtain a final concentration of 0.1% SDS. Diluted samples were centrifuged at 13,000 rpm for 10 min. The supernatant was pre-cleared with protein A/G or Dynabeads M-280 Streptavidin beads (ThermoFisher) and immunoprecipitated for 3–12 h using 3 µg antibodies and 40 µl protein A/G or Dynabeads M-280 Streptavidin beads. The beads were washed twice with high salt wash buffer A (pH 7.9, 50 mM Hepes, 500 mM NaCl, 1 mM EDTA, 1% Triton, 0.1% Sodium deoxycholate, and 0.1% SDS), and once with wash buffer B (pH 7.9, 50 mM Hepes, 250 mM LiCl, 1 mM EDTA, 1% Triton, 0.1% Sodium deoxycholate, 0.5% NP-40). The bound chromatin fragments were eluted with elution buffer (pH 8.0, 50 mM Tris, 10 mM EDTA, 1% SDS) for 5 min at 65 °C. Eluted DNA-proteins complexes were treated with RNase A and crosslinks were reversed overnight at 65 °C. Proteinase K was then added to digest proteins for 1 h at 55 °C. DNA was further purified using PCR Purification Kit (QIAGEN) and analysed by PCR on a QuantStudio 7 Flex Real Time PCR System (Applied Biosystem). PCR parameters were 95 °C for 2 min and 40 cycles of 95 °C for 15 s, 60 °C for 15 s, and 72 °C for 15 s, followed by 72 °C for 1 min. All the ChIP-qPCR data presented were at least three biological replicates. Primer sequences are in Supplementary Table 4. Error bars represent standard deviation (three biological replicates).

For ChIP-seq, sequencing library was constructed using TruSeq DNA sample Prep Kits (Illumina) and adapter dimers were removed by agarose gels electrophoresis. Sized selected and purified DNA libraries were sequenced on an Illumina Hiseq 2500 machine (Bauer core facility at Harvard University) to obtain 50 bp single-ended reads. ChIP-seq reads were quality controlled with fastqc (v0.11.5) and mapped to the human genome reference (GRCh37/hg19) using bowtie2 (v2.2.9) with default parameters or bowtie (v1.2.2) with parameters -v2 -k1-best. Bam files were generated with samtools 1.3.1, which was followed by making bigwig files with deeptools (v3.0.2) (binsize 10). Reads were normalized to Reads Per Genome Coverage (RPGC) with deeptools (v3.0.2) bamCoverage function. To analyse read density at TSS regions, we made heatmaps and metaplots of ChIP-seq samples. TSS was centered in the regions plotted and data were tabulated with the same distance relative to TSS. Matrix files were generated using computematrix function of deeptools (v3.0.2). Based on generated matrix file, heatmaps were generated by PlotHeatmap function, and profiles were generated by plotprofile function or in Prism.

To analyse read density and correlation between different ChIP-seq samples, we performed Spearman correlation analysis. Reads density was analysed at all hg19 annotated TSSs ( $n = 56,335$ ) with multiBigwigSummary function from deeptools (v3.0.2) to get a npz matrix file. The heatmap Spearman or Pearson correlation was generated by plotCorrelation function of deeptools (v3.0.2). The heatmaps generated in this study also included all annotated human genes (hg19). The gene list was obtained from <https://genome.ucsc.edu>. Promoter regions were defined as  $\pm 2$  kb from TSSs. Peak overlaps were analysed by bedtools (v3.0.2) intersect function.

For co-occupancy analysis in Extended Data Fig. 2, peak calling of TEX10, H2AK119ub1, and H3K9me3 was performed with MACS2 (2.1.1.20160309) with Input ChIP-seq sample as control (-p 0.05–broad, –broad-cutoff 0.05, FoldChang>2.5, Length>800 bp).

For defining TEX10-bound targets in Fig. 2, TEX10 peaks were called using HOMER (version 4.9) with the -style histone option and siTEX10 ChIP-seq as background. TEX10-bound genes were defined as those that had 50 or more TEX10 reads in the TSS  $\pm$  1 kb region ( $n = 7,827$ ); all others were considered unbound ( $n = 13,177$ ).

For defining Polycomb target genes in Figs. 2, 3, H2AK119ub1 ChIP-seq data from HEK 293FT cells were used. DeepTools was used to count reads in TSS  $\pm$  2 kb regions. K-means clustering was performed with  $k = 2$ . Cluster one was H2AK119ub1 enriched and counted as Polycomb target genes. Venn diagrams in Extended Data Fig. 8 were made based on the number of overlapping target genes. DeepTools was used to count reads in TSS  $\pm$  2 kb regions. K-means clustering was performed with a fixed value of  $k = 3$ . Cluster one was counted as target genes.

The sources of ChIP-seq data used in this study are listed in Supplementary Table 5.

## Statistical tests

For RNA-seq, PRO-seq and ChIP-seq, statistical significance for comparisons was assessed by Wilcoxon (unpaired) or Mann–Whitney (pairwise) tests. The test used and error bars are defined in each figure legend.

Significance for immunostaining foci was evaluated using unpaired two-tail student's *t*-test. All the RT-qPCR and ChIP-qPCR data are represented as mean  $\pm$  s.d. using GraphPad Prism 8 software. Volcano plots of Mass spec results were made with Microsoft Excel.

## Reporting summary

Further information on research design is available in the [Nature Research Reporting Summary](#) linked to this paper.

## Data availability

The raw mass spectrometry data were deposited with accession number [PXD027966](#) and [PXD029403](#). The raw and processed high-throughput sequencing data have been deposited at NCBI Gene Expression Omnibus under accession [GSE175678](#). [Source data](#) are provided with this paper.

## Code availability

Software and algorithms used in this study are listed in [Supplementary Table 6](#).

## References

1. Schuettengruber, B., Bourbon, H. M., Di Croce, L. & Cavalli, G. Genome regulation by Polycomb and Trithorax: 70 years and counting. *Cell* **171**, 34–57 (2017).
2. Simon, J. A. & Kingston, R. E. Occupying chromatin: Polycomb mechanisms for getting to genomic targets, stopping transcriptional traffic, and staying put. *Mol. Cell* **49**, 808–824 (2013).
3. Margueron, R. & Reinberg, D. The Polycomb complex PRC2 and its mark in life. *Nature* **469**, 343–349 (2011).
4. Shipkovenska, G., Durango, A., Kalocsay, M., Gygi, S. P. & Moazed, D. A conserved RNA degradation complex required for spreading and epigenetic inheritance of heterochromatin. *eLife* **9**, e54341 (2020).
5. Holoch, D. & Moazed, D. RNA-mediated epigenetic regulation of gene expression. *Nat. Rev. Genet.* **16**, 71–84 (2015).
6. Fromm, L. et al. Reconstitution of the complete pathway of ITS2 processing at the pre-ribosome. *Nat. Commun.* **8**, 1787 (2017).
7. Shao, Z. et al. Stabilization of chromatin structure by PRC1, a Polycomb complex. *Cell* **98**, 37–46 (1999).
8. de Napoles, M. et al. Polycomb group proteins Ring1A/B link ubiquitylation of histone H2A to heritable gene silencing and X inactivation. *Dev. Cell* **7**, 663–676 (2004).
9. Wang, H. et al. Role of histone H2A ubiquitination in Polycomb silencing. *Nature* **431**, 873–878 (2004).
10. Muller, J. et al. Histone methyltransferase activity of a *Drosophila* Polycomb group repressor complex. *Cell* **111**, 197–208 (2002).
11. Cao, R. et al. Role of histone H3 lysine 27 methylation in Polycomb-group silencing. *Science* **298**, 1039–1043 (2002).

12. Czernin, B. et al. *Drosophila* enhancer of Zeste/ESC complexes have a histone H3 methyltransferase activity that marks chromosomal Polycomb sites. *Cell* **111**, 185–196 (2002).
13. Kuzmichev, A., Nishioka, K., Erdjument-Bromage, H., Tempst, P. & Reinberg, D. Histone methyltransferase activity associated with a human multiprotein complex containing the Enhancer of Zeste protein. *Genes Dev.* **16**, 2893–2905 (2002).
14. Gao, Z. et al. PCGF homologs, CBX proteins, and RYBP define functionally distinct PRC1 family complexes. *Mol. Cell* **45**, 344–356 (2012).
15. Margueron, R. et al. Role of the Polycomb protein EED in the propagation of repressive histone marks. *Nature* **461**, 762–767 (2009).
16. Zhao, J. et al. RYBP/YAF2–PRC1 complexes and histone H1-dependent chromatin compaction mediate propagation of H2AK119ub1 during cell division. *Nat. Cell Biol.* **22**, 439–452 (2020).
17. Kalb, R. et al. Histone H2A monoubiquitination promotes histone H3 methylation in Polycomb repression. *Nat. Struct. Mol. Biol.* **21**, 569–571 (2014).
18. Kasinath, V. et al. JARID2 and AEBP2 regulate PRC2 in the presence of H2AK119ub1 and other histone modifications. *Science* **371**, eabc3393 (2021).
19. Blackledge, N. P. et al. Variant PRC1 complex-dependent H2A ubiquitylation drives PRC2 recruitment and polycomb domain formation. *Cell* **157**, 1445–1459 (2014).
20. Blackledge, N. P. et al. PRC1 catalytic activity is central to Polycomb system function. *Mol. Cell* **77**, 857–874.e859 (2020).
21. Tamburri, S. et al. Histone H2AK119 mono-ubiquitination is essential for Polycomb-mediated transcriptional repression. *Mol. Cell* **77**, 840–856.e845 (2020).
22. Grau, D. J. et al. Compaction of chromatin by diverse Polycomb group proteins requires localized regions of high charge. *Genes Dev.* **25**, 2210–2221 (2011).
23. Eskeland, R. et al. Ring1B compacts chromatin structure and represses gene expression independent of histone ubiquitination. *Mol. Cell* **38**, 452–464 (2010).
24. Francis, N. J., Kingston, R. E. & Woodcock, C. L. Chromatin compaction by a Polycomb group protein complex. *Science* **306**, 1574–1577 (2004).

25. Margueron, R. et al. Ezh1 and Ezh2 maintain repressive chromatin through different mechanisms. *Mol. Cell* **32**, 503–518 (2008).
26. Terranova, R. et al. Polycomb group proteins Ezh2 and Rnf2 direct genomic contraction and imprinted repression in early mouse embryos. *Dev. Cell* **15**, 668–679 (2008).
27. Lau, M. S. et al. Mutation of a nucleosome compaction region disrupts Polycomb-mediated axial patterning. *Science* **355**, 1081–1084 (2017).
28. Plys, A. J. et al. Phase separation of Polycomb-repressive complex 1 is governed by a charged disordered region of CBX2. *Genes Dev.* **33**, 799–813 (2019).
29. Tatavosian, R. et al. Nuclear condensates of the Polycomb protein chromobox 2 (CBX2) assemble through phase separation. *J. Biol. Chem.* **294**, 1451–1463 (2019).
30. Fursova, N. A. et al. Synergy between variant PRC1 complexes defines Polycomb-mediated gene repression. *Mol. Cell* **74**, 1020–1036.e1028 (2019).
31. Scelfo, A. et al. Functional landscape of PCGF proteins reveals both RING1A/B-dependent-and RING1A/B-independent-specific activities. *Mol. Cell* **74**, 1037–1052.e1037 (2019).
32. Breiling, A., Turner, B. M., Bianchi, M. E. & Orlando, V. General transcription factors bind promoters repressed by Polycomb group proteins. *Nature* **412**, 651–655 (2001).
33. Dellino, G. I. et al. Polycomb silencing blocks transcription initiation. *Mol. Cell* **13**, 887–893 (2004).
34. Iglesias, N. et al. Native chromatin proteomics reveals a role for specific nucleoporins in heterochromatin organization and maintenance. *Mol. Cell* **77**, 51–66.e58 (2020).
35. Zhao, W. et al. Polycomb group RING finger proteins 3/5 activate transcription via an interaction with the pluripotency factor Tex10 in embryonic stem cells. *J. Biol. Chem.* **292**, 21527–21537 (2017).
36. Huttlin, E. L. et al. Dual proteome-scale networks reveal cell-specific remodeling of the human interactome. *Cell* **184**, 3022–3040.e3028 (2021).
37. Fanis, P. et al. Five friends of methylated chromatin target of protein-arginine-methyltransferase[prmt]-1 (chtop), a complex linking arginine methylation to

- desumoylation. *Mol. Cell Proteomics* **11**, 1263–1273 (2012).
38. Healy, E. et al. PRC2.1 and PRC2.2 synergize to coordinate H3K27 trimethylation. *Mol. Cell* **76**, 437–452.e436 (2019).
39. Isono, K. et al. SAM domain polymerization links subnuclear clustering of PRC1 to gene silencing. *Dev. Cell* **26**, 565–577 (2013).
40. Cheutin, T. & Cavalli, G. Progressive Polycomb assembly on H3K27me3 compartments generates Polycomb bodies with developmentally regulated motion. *PLoS Genet.* **8**, e1002465 (2012).
41. Castle, C. D., Cassimere, E. K. & Denicourt, C. LAS1L interacts with the mammalian Rix1 complex to regulate ribosome biogenesis. *Mol. Biol. Cell* **23**, 716–728 (2012).
42. Kwak, H., Fuda, N. J., Core, L. J. & Lis, J. T. Precise maps of RNA polymerase reveal how promoters direct initiation and pausing. *Science* **339**, 950–953 (2013).
43. Gasse, L., Flemming, D. & Hurt, E. Coordinated ribosomal ITS2 RNA processing by the Las1 complex integrating endonuclease, polynucleotide kinase, and exonuclease activities. *Mol. Cell* **60**, 808–815 (2015).
44. Pillon, M. C., Sobhany, M., Borgnia, M. J., Williams, J. G. & Stanley, R. E. Grc3 programs the essential endoribonuclease Las1 for specific RNA cleavage. *Proc. Natl Acad. Sci. USA* **114**, E5530–E5538 (2017).
45. Heindl, K. & Martinez, J. Nol9 is a novel polynucleotide 5'-kinase involved in ribosomal RNA processing. *EMBO J.* **29**, 4161–4171 (2010).
46. Sanso, M. et al. P-TEFb regulation of transcription termination factor Xrn2 revealed by a chemical genetic screen for Cdk9 substrates. *Genes Dev.* **30**, 117–131 (2016).
47. Buhler, M., Haas, W., Gygi, S. P. & Moazed, D. RNAi-dependent and -independent RNA turnover mechanisms contribute to heterochromatic gene silencing. *Cell* **129**, 707–721 (2007).
48. Mattout, A. et al. LSM2-8 and XRN-2 contribute to the silencing of H3K27me3-marked genes through targeted RNA decay. *Nat. Cell Biol.* **22**, 579–590 (2020).
49. Davidovich, C., Zheng, L., Goodrich, K. J. & Cech, T. R. Promiscuous RNA binding by Polycomb repressive complex 2. *Nat. Struct. Mol. Biol.* **20**, 1250–1257 (2013).

50. Kaneko, S., Son, J., Shen, S. S., Reinberg, D. & Bonasio, R. PRC2 binds active promoters and contacts nascent RNAs in embryonic stem cells. *Nat. Struct. Mol. Biol.* **20**, 1258–1264 (2013).
51. Beltran, M. et al. The interaction of PRC2 with RNA or chromatin is mutually antagonistic. *Genome Res.* **26**, 896–907 (2016).
52. Wang, X. et al. Molecular analysis of PRC2 recruitment to DNA in chromatin and its inhibition by RNA. *Nat. Struct. Mol. Biol.* **24**, 1028–1038 (2017).
53. Long, Y. et al. RNA is essential for PRC2 chromatin occupancy and function in human pluripotent stem cells. *Nat. Genet.* **52**, 931–938 (2020).
54. Chen, G. et al. Chemically defined conditions for human iPSC derivation and culture. *Nat. Methods* **8**, 424–429 (2011).
55. Navarrete-Perea, J., Yu, Q., Gygi, S. P. & Paulo, J. A. Streamlined tandem mass tag (SL-TMT) protocol: an efficient strategy for quantitative (phospho)proteome profiling using tandem mass tag-synchronous precursor selection-MS3. *J. Proteome Res.* **17**, 2226–2236 (2018).
56. Reimer, K. A., Mimoso, C. A., Adelman, K. & Neugebauer, K. M. Co-transcriptional splicing regulates 3' end cleavage during mammalian erythropoiesis. *Mol. Cell* **81**, 998–1012.e1017 (2021).
57. Nowak, D. E., Tian, B. & Brasier, A. R. Two-step cross-linking method for identification of NF-κB gene network by chromatin immunoprecipitation. *Biotechniques* **39**, 715–725 (2005).

## Acknowledgements

We thank the Nikon Imaging Center at the Department of Cell Biology at Harvard Medical School for access to microscopes and advice, the HMS Nascent Transcriptomics Core for PRO-seq, members of the Moazed laboratory for helpful discussions, and K. D'Orazio, S. Parhad, H. Saini, A. Tatarakis and J. Yu for comments on the manuscript. This work was supported by the National Institutes of Health, grants R01GM134539 to K.A. and RO1GM072805 to D.M. D.M. is an investigator of the Howard Hughes Medical Institute.

## Author information

### Affiliations

1. Howard Hughes Medical Institute, Department of Cell Biology, Blavatnik Institute, Harvard Medical School, Boston, MA, USA

Haining Zhou, Tiasha A. Shafiq, Gergana Shipkovenska & Danesh Moazed

2. Department of Biological Chemistry and Molecular Pharmacology, Blavatnik Institute, Harvard Medical School, Boston, MA, USA

Chad B. Stein & Karen Adelman

3. Department of Cell Biology, Blavatnik Institute, Harvard Medical School, Boston, MA, USA

Marian Kalocsay, Joao A. Paulo & Steven P. Gygi

4. Initiative for Genome Editing and Neurodegeneration, Department of Cell Biology, Blavatnik Institute, Harvard Medical School, Boston, MA, USA

Jiuchun Zhang

5. Precision Medicine Institute, the First Affiliated Hospital, Sun Yat-sen University, Guangzhou, China

Zhenhua Luo

## Contributions

H.Z. and D.M. conceived the study and designed experiments. H.Z. carried out mutagenesis, cell line construction, protein purification, ChIP-seq, RNA-seq, ChIP-qPCR, RT-qPCR, Y2H assays, bioinformatics, microscopy and data analysis. C.B.S. prepared permeabilized cells for PRO-seq and C.B.S. and K.A. analysed ChIP-seq and PRO-seq data. T.A.S. constructed 5xtetO-H2B-CITRINE reporter cell lines and contributed to experimental design and data analysis. G.S. provided information on the fission yeast rixosome prior to publication. M.K. and J.A.P. performed mass spectrometry. J.Z. generated knockout cell lines. Z.L. performed bioinformatic analysis of RNA-seq data. S.P.G., K.A. and D.M. supervised research. H.Z. and D.M. wrote the original draft of the manuscript. C.B.S. and K.A. contributed to editing and writing subsequent drafts. All authors discussed the results and commented on the manuscript.

## Corresponding author

Correspondence to [Danesh Moazed](#).

# Ethics declarations

## Competing interests

The authors declare no competing interests.

## Peer review

## Peer review information

*Nature* thanks Robert Fisher and the other, anonymous, reviewers for their contribution to the peer review of this work.

## Additional information

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## Extended data figures and tables

### [Extended Data Fig. 1 Physical association of the rixosome and Polycomb complexes.](#)

**a**, Diagrams showing the composition of the rixosome and Polycomb complexes. **b**, Silver-stained gel of Flag immunoprecipitations from untagged (no tag) and Flag-WDR18 HEK293FT cells. **c**, Volcano plot of TMT mass spectrometry results showing log2-fold changes in proteins enrichment in Flag immunoprecipitations from Flag-WDR18 versus untagged cells from two independent experiments. *p* values calculated by two-sided student's t test. Rixosome, PRC1, PCR2, and HP1 proteins are highlighted in green, blue, magenta, and yellow, respectively. **d**, Immunoprecipitations (IP) showing the association of RING1B with rixosome subunits PELP1, TEX10, SENP3 and Flag-WDR18 in HEK293FT cells. **e**, Volcano plot of mass spectrometry results showing log2-fold changes in proteins enrichment in Flag immunoprecipitations from Flag-CBX4 versus untagged cells from two independent experiments. *p* values calculated by two-sided student's t test. Rixosome and PRC1 subunits are highlighted in green and blue, respectively. **f**, Yeast two-hybrid assays. Yeast cells transformed with the indicated plasmids were plated onto double dropout (Non-selective) (SC-Trp, -Leu) or quadruple dropout (Selective)(SC-Trp, -Leu, -His, -Ade) medium. AD, Activation Domain; BD, Binding Domain. **g**, Diagram of the RING1B protein and the binding activities of its indicated truncations. CC1 and CC2,

coiled-coil domains 1 and 2. **h, i**, Pull-down assays using bacterially expressed and purified TEX10 proteins and the indicated bead-immobilized GST or GST-fusion RING1B WT or mutant proteins. TEX10 detected by immunoblotting using an anti-TEX10 antibody. GST-tagged proteins were stained with Coomassie. The assays were performed three time independently with similar results.

### Extended Data Fig. 2 The rixosome is preferentially enriched at promoter regions.

**a, b**, Immunoblot validation of siRNA-mediated TEX10 (**a**) and MDN1 (**b**) knockdowns. **c, d**, Average distribution of the indicated ChIP-seq reads at the TEX10-bound genes ( $n = 7,827$ ) versus TEX10-unbound genes ( $n = 13,177$ ). Read counts per gene were averaged in 50-nt bins, using summed reads in the window  $\pm 1\text{kb}$  from TSS. **e**, Average distribution of TEX10 and MDN1 ChIP-seq signal is shown relative to all annotated (hg19) transcription start sites (TSS), transcription termination sites (TES), gene bodies, and enhancers. Enrichment levels ( $\log_2$ ) were normalized with Reads Per Genome Coverage. Read counts per gene were summed in 50-nt bins. **f**, Average distribution of H2AK119ub1, H3K27me3, and H3K9me3 ChIP-seq signal is shown relative to all annotated (hg19) transcription start sites (TSS), transcription termination sites (TES), gene bodies, and enhancers. Enrichment levels ( $\log_2$ ) were normalized with Reads Per Genome Coverage. Read counts per gene were summed in 50-nt bins. **g**, Heatmap representations of ChIP-seq of TEX10, MDN1, RING1B and histone modifications H2AK119ub1 and H3K9me3 at TEX10 peak regions. Rank order is from most to least TEX10 signal. Enrichment levels ( $\log_2$ ) were normalized with Reads Per Genome Coverage. Read counts per peak region were summed in 50-nt bins. **h**, Heatmap representations of ChIP-seq of TEX10, MDN1, RING1B and histone modifications H2AK119ub1 and H3K9me3 at H2AK119ub1 peak regions. Rank order is from most to least TEX10 signal. Enrichment levels ( $\log_2$ ) were normalized with Reads Per Genome Coverage. Read counts per peak region were summed in 50-nt bins. **i**, Heatmap representations of ChIP-seq of TEX10, MDN1, RING1B and histone modifications H2AK119ub1 and H3K9me3 at H3K9me3 peak regions. Rank order is from most to least H3K9me3 signal. Enrichment levels ( $\log_2$ ) were normalized with Reads Per Genome Coverage. Read counts per peak region were summed in 50-nt bins. **j**, Matrix depicting Spearman correlation coefficients between ChIP-seq datasets in HEK293 cells, calculated using read counts in all the genomic loci from **e–g**.

### Extended Data Fig. 3 Colocalization of the rixosome with Polycomb bodies.

**a, b**, Immunoblot validation of siRNA-mediated WDR18 (**a**) and EZH2 (**b**) knockdowns. **c**, Validation of siRNA knockdowns (48 h after transfection) of rixosome

subunits, RING1B, and EZH2. **d**, Immunofluorescence colocalization of rixosome subunits MDN1 with EZH2-stained Polycomb bodies in cells treated with the indicated siRNA. DNA was stained with DAPI (blue). Scale bar, 5  $\mu$ m. **e**, Quantification of MDN1 and EZH2 fluorescence intensity in **d**. *p* values are from two-sided student's t-tests. Data are presented as mean values +/- SEM. **f**, Immunofluorescence of MDN1 (green), the nucleolar NPM1 protein (purple), and EZH2-stained foci (yellow). DNA was stained with DAPI (blue). **g**, Quantification of overlap between MDN1 foci and EZH2 or NPM1 per nucleus in the wild type cells. *p* values are from two-sided student's t-tests. Data are presented as mean values +/- SEM.

[Source data](#)

#### [Extended Data Fig. 4 Polycomb body-enrichment of rixosome is EZH1/2- and RING1A/B-dependent.](#)

**a**, Immunofluorescence experiments showing colocalization of rixosome subunits MDN1 and WDR18 with the nucleolar marker NPM1 in wild-type (WT) and EZH1/2 double knockout cells, and with NPM1 and EZH2 in RING1A/B double knockout cells. Scale bar, 5  $\mu$ m. **b**, **c**, Quantification of change in MDN1 (**b**) or NPM1 (**c**) foci per nucleus in the indicated cells. *p* values are from two-sided student's t-tests. Data are presented as mean values +/- SEM.

[Source data](#)

#### [Extended Data Fig. 5 Rixosome effects on H3K27me3 and H2AK119ub1, and rixosome-Polycomb association.](#)

**a**, Co-fractionation of rixosome and PRC1 subunits. Flag-NOL9-associated and PHC2–Flag-associated proteins purified from cells with wild-type or *RING1B-2A* and were subjected to 10–30% sucrose gradient sedimentation. Fractions were collected and adsorbed to Strataclean beads and analyzed by immunoblotting with the indicated antibodies. **b**, Average distribution of H2AK119ub1 ChIP-seq reads (log2) for all annotated genes in wild type (WT) and *RING1B-2A* (2A) mutant HEK293FT cells. Enrichment levels were normalized with Reads Per Genome Coverage. Read counts per gene were summed in 50-nt bins. **c**, ChIP-qPCR experiments showing enrichment of H2AK119ub1 at the indicated target genes in siCtrl, siNOL9, and siRING1B treated HEK293FT cells. Error bars represent standard deviations for three biological replicates. Data are presented as mean values +/- SEM. **d**, ChIP-qPCR showing enrichment of H3K27me3 at the indicated target genes in siCtrl, siNOL9, and siEZH2 treated HEK293FT cells. Error bars represent standard deviations for three biological replicates. Data are presented as mean values +/- SEM. **e**, Heatmap representations of

H2AK119ub1 ChIP-seq from control cells compared to cells depleted of NOL9 at TSS flanking regions. Rank order is by H2AK119ub1 signal from siCtrl cells. Enrichment levels ( $\log_2$ ) were normalized with Reads Per Genome Coverage. Color bar at bottom indicates range of read counts per 50-nt bin. **f**, Heatmap representations of H3K27me3 ChIP-seq from control cells compared to cells depleted of NOL9 at TSS flanking regions. Rank order is by H3K27me3 signal from siCtrl cells. Enrichment levels ( $\log_2$ ) were normalized with Reads Per Genome Coverage. Color bar at bottom indicates range of read counts per 50-nt bin.

[Source data](#)

**Extended Data Fig. 6 Rixosome subunits, H2AK119ub1, RING1B, and H3K27me3 are preferentially enriched at PRO-seq siNOL9-upregulated genes.**

**a**, Growth curves show cell number changes at indicated time points after knockdowns with siCtrl, siSUV39H1, siEZH2, siNOL9, or siLAS1L in HEK293FT cells. Error bars represent standard deviation for three biological replicates. Data are presented as mean values  $\pm$  SEM. **b**, Average distribution of PRO-seq signal is shown at genes downregulated by siNOL9 ( $N = 30$ ). Data are shown in 25-nt bins. **c**, Average distribution of PRO-seq signal is shown at a set of genes unaffected by siNOL9 ( $N = 30$ ) which were expression matched for the downregulated genes in **b**. Data are shown in 25-nt bins. **d**, Violin plots depict the  $\log_2$  (fold change) in PRO-seq for siNOL9 downregulated ( $N = 30$ ) and unaffected ( $N = 30$ ) genes in siNOL9, siRING1B, *RING1AB* DKO, and *EED* KO cells. Knockout cells were treated with control siRNA. *p*-values are from two-tailed Mann-Whitney test.  $P = 0.3581$  for siRING1AB,  $P = 0.6438$  for *RING1AB* DKO,  $P = 0.6228$  for *EED* KO. **e**, Violin plots showing read counts for the indicated ChIP-seq experiments. Reads in were summed  $\pm 1$  kb from TSSs for the gene groups indicated. Violin plots depict the range of values, with median indicated by a line. *p*-values are from two-tailed Mann-Whitney test.  $P = 0.0034$  for *TEX10*,  $P = 0.0648$  for *MDN1*,  $P = 0.0058$  for *RING1B*,  $P = 0.017$  for *H2AK119ub1*,  $P = 0.0028$  for *H3K27me3*,  $P = 0.0276$  for *H3K9me3*. n.s., not significant. **f–k**, Average distribution of the indicated ChIP-seq reads at siNOL9-upregulated or siNOL9-unaffected genes. Read counts per gene were summed in 50-nt bins. **l–m**, Venn diagrams showing the overlap between siNOL9-upregulated (**l**) and siNOL9-downregulated (**m**) genes with genes upregulated in *RING1AB* DKO or *EED* KO cells in PRO-seq experiments.

[Source data](#)

**Extended Data Fig. 7 Coregulation of target genes by the rixosome and PRCs.**

**a–c**, Average distribution of the indicated ChIP-seq reads ( $\log_2$ ) for genes upregulated and downregulated in *LASIL KD* (a) and *RING1B-2A*(b), and upregulated genes in *EED KO* and *RING1A/B DKO* (c) RNA-seq experiments from HEK293FT cells. Enrichment levels were normalized with Reads Per Genome Coverage. Read counts per gene were summed in 50-nt bins. **d**, Dot plots showing RNA-seq changes in the expression of siNOL9-upregulated or siNOL9-downregulated genes in PRO-seq experiments in HEK293FT cells. siNOL9 PRO-seq upregulated genes have increased RNA levels in siNOL9, *EED KO*, and *RING1A/B DKO* cells.  $p$  value is from the two-tailed Wilcoxon test. The measure of center is median. **e**, RNA-seq experiments showing increased *HOX* gene expression in siNOL9 and *RING1A/B DKO* but not *EED KO* cells.  $P$  value is from the two-tailed Wilcoxon test. The measure of center is median. **f**, Genomic snapshots of RNA-seq reads showing the effect of siRNA knockdown of NOL9 and *RING1A/B DKO* on the expression of the indicated genes in HEK293FT cells. **g, h**, RT-qPCR assays showing that siRNA knockdown of rixosome subunits results in increased expression of *PCDH10* in wild-type (WT), but not *EZH1/2 DKO* or *RING1A/B DKO* HEK293FT cells. Actin (*ACTB*) served as a normalization control. Every knockdown was normalized to siCtrl. Nucleolar PES1 and NPM1 served as controls for possible non-specific effects resulting nucleolar perturbations. Error bars represent standard deviations for three biological replicates. Data are presented as mean values  $\pm$  SEM. **i**, RT-qPCR experiments showing the effect of the indicated siRNA knockdowns on the indicated Polycomb and rixosome target genes in wild-type (WT) cells and *RING1A/B DKO* cells. Actin (*ACTB*) served as a normalization control. Every knockdown was normalized to siCtrl. Error bars represent standard deviations for three biological replicates. Data are presented as mean values  $\pm$  SEM.

[Source data](#)

### **Extended Data Fig. 8 Colocalization of the rixosome with H2AK119ub1 in human ES and HeLa cells.**

**a**, Matrix depicting Spearman correlation coefficients between ChIP-seq datasets calculated using read counts summed  $\pm 2$  kb for all annotated gene TSSs (hg19) in human embryonic stem (ES) cells. **b**, Heatmap representations of ChIP-seq of TEX10 and histone modifications (H2AK119ub1, H3K27me3, H3K9me3, H3K79me3, and H3K4me3,) in human ES cells. Rank order is from most to least TEX10 signal. Enrichment levels ( $\log_2$ ) were normalized with Reads Per Genome Coverage. Read counts per gene were summed in 50-nt bins. **c**, Matrix depicting Spearman correlation coefficients between ChIP-seq datasets calculated using read counts summed  $\pm 2$  kb for all annotated gene TSSs (hg19) in HeLa cells. **d**, Heatmap representations of ChIP-seq of MDN1, H2AK119ub1, and H3K27me3 in HeLa cells. Rank order is from most to least TEX10 signal. Enrichment levels ( $\log_2$ ) were normalized with Reads Per

Genome Coverage. Read counts per gene were summed in 50-nt bins. **e**, Genomic snapshots of ChIP-seq reads at Polycomb target *HOXA* cluster in human ES (top) and HeLa (bottom) cells for the indicated rixosome subunits or Polycomb histone modifications. Enrichment levels ( $\log_2$ ) were normalized with Reads Per Genome Coverage. **f**, Venn diagram showing overlap between TEX10- and H2AK119ub1-enriched TSSs in human ES cells. Hypergeometric probability  $p$  values, 2.9e-4262 **g**, Venn diagram showing overlap between MDN1- and H2AK119ub1-enriched TSSs in HeLa cells. Hypergeometric probability  $p$  values, 8.9e-2935.

[Source data](#)

**Extended Data Fig. 9 Role of the rixosome in Polycomb silencing is not cell type specific.**

**a**, Venn diagram showing overlap among genes upregulated in RNA-seq analysis of siNOL9, siLAS1L, and siTEX10 HeLa cells. Hypergeometric probability  $p$  values: siNOL9 vs siLAS1L, 2.6e-765; siNOL9 vs siTEX10, 1.4e-859; siLAS1L vs siTEX10, 1.6e-853. **b**, Same as in **a** but showing overlap among genes upregulated in siRING1B in *RING1A* KO (siRING1B, *RING1A*-/-), siNOL9 (in wild type), and siEZH2 (in wild type). Hypergeometric probability  $p$  values: siNOL9 vs siRING1B, *RING1A*-/-, 4.1e-441; siNOL9 vs siEZH2, 4.6e-411; siEZH2 vs siRING1B, *RING1A*-/-, 2.9e-284. **c**, Same as in **a** but showing overlap among genes upregulated in siRING1B, *RING1A*-/-, siLAS1L (in wild type), and siEZH2 (in wild type). Hypergeometric probability  $p$  values: siLAS1L vs siRING1B, *RING1A*-/-, 1.4e-496; siLAS1L vs siEZH2, 1.7e-298. **d**, Same as in **a** but showing overlap among genes upregulated in siRING1B, *RING1A*-/-, siTEX10 (in wild type), and siEZH2 (in wild type). Hypergeometric probability  $p$  values: siTEX10 vs siRING1B, *RING1A*-/-, 1.8e-391; siTEX10 vs siEZH2, 9.1e-347. **e**, Table showing the percentages of overlapping upregulated genes between rixosome and PRC depletions in panels **a-d**. **f**, Dot plots of RNA-seq experiments showing changes in the expression of 39 *HOX* genes in HeLa cells.  $p$  values are from two-tailed Wilcoxon test. The measure of center is median. **g**, Average distribution of indicated ChIP-seq reads ( $\log_2$ ) for genes upregulated by siRNA depletion of TEX10, LAS1L, NOL9, and RING1B (*RING1A*-/-) in HeLa cell RNA-seq experiments. Enrichment levels were normalized with Reads Per Genome Coverage. Read counts per gene were summed in 50-nt bins. **h**, Same as in **g** but showing siNOL9 downregulated genes.

[Source data](#)

**Extended Data Fig. 10 NOL9, LAS1L, and XRN2 catalytic point mutations.**

**a**, Immunoblot showing protein levels in control siControl (siCtrl), siNOL9, siNOL9+NOL9 wild type expressing plasmid, and siNOL9+ NOL9-K312A expressing plasmids. Actin served as a loading control. **b**, Immunoblot showing protein levels in control siControl (siCtrl), siLAS1L, siLAS1L+LAS1L wild type expressing plasmid, and siLAS1L+ LAS1L-R155A/H160A (LAS1l-2A) expressing plasmids. GAPDH served as a loading control. **c**, Immunoblot showing protein levels in control siControl (siCtrl), siXRN2-1 (siRNA 1), siXRN2-2 (siRNA 2), siXRN2-1+XRN2 wild type expressing plasmid, and siXRN2-1+ XRN2 E203G expressing plasmids. GAPDH served as a loading control. **d**, RT-qPCR analysis of RNA levels of H2B-CTRN in the indicated siRNA-treated and NOL9-rescued HEK293FT cells after 21 days of Doxycycline treatment. RNA expression levels were normalized to *ACTB*, and every knockdown was normalized to siCtrl. Error bars represent standard deviations for three biological replicates. Data are presented as mean values  $\pm$  SEM.

[Source data](#)

## Supplementary information

### [Supplementary Information](#)

This file contains Supplementary Tables 1–6 and Supplementary Fig. 1, which includes the uncropped gels and blots for the main and Extended Data Figures.

### [Reporting Summary](#)

### [Supplementary Table 7](#)

The results of mass spectrometry identification of proteins that co-purify with Flag–WDR18 and Flag–NOL9.

### [Supplementary Table 8](#)

The results of mass spectrometry identification of proteins that co-purify with Flag–PHC2 and Flag–CBX4.

## Source data

### [Source Data Fig. 3](#)

### [Source Data Fig. 4](#)

[Source Data Fig. 5](#)

[Source Data Extended Data Fig. 3](#)

[Source Data Extended Data Fig. 4](#)

[Source Data Extended Data Fig. 5](#)

[Source Data Extended Data Fig. 6](#)

[Source Data Extended Data Fig. 7](#)

[Source Data Extended Data Fig. 8](#)

[Source Data Extended Data Fig. 9](#)

[Source Data Extended Data Fig. 10](#)

## Rights and permissions

**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

Zhou, H., Stein, C.B., Shafiq, T.A. *et al.* Rixosomal RNA degradation contributes to silencing of Polycomb target genes. *Nature* **604**, 167–174 (2022).  
<https://doi.org/10.1038/s41586-022-04598-0>

- Received: 22 June 2021
- Accepted: 28 February 2022
- Published: 30 March 2022
- Issue Date: 07 April 2022
- DOI: <https://doi.org/10.1038/s41586-022-04598-0>

## Share this article

Anyone you share the following link with will be able to read this content:

[Get shareable link](#)

Sorry, a shareable link is not currently available for this article.

[Copy to clipboard](#)

Provided by the Springer Nature SharedIt content-sharing initiative

## Further reading

- [\*\*An added layer of repression for human genes\*\*](#)

- Michael Uckelmann
- Chen Davidovich

*Nature* (2022)

## [\*\*An added layer of repression for human genes\*\*](#)

- Michael Uckelmann
- Chen Davidovich

News & Views 30 Mar 2022

---

This article was downloaded by **calibre** from <https://www.nature.com/articles/s41586-022-04598-0>

- Article
- [Published: 06 April 2022](#)

# Mapping the energetic and allosteric landscapes of protein binding domains

- [Andre J. Faure](#) ORCID: [orcid.org/0000-0002-4471-5994](https://orcid.org/0000-0002-4471-5994)<sup>1 na1</sup>,
- [Júlia Domingo](#) ORCID: [orcid.org/0000-0001-6359-2785](https://orcid.org/0000-0001-6359-2785)<sup>1,4 na1</sup>,
- [Jörn M. Schmiedel](#) ORCID: [orcid.org/0000-0002-6842-9837](https://orcid.org/0000-0002-6842-9837)<sup>1 na1</sup>,
- [Cristina Hidalgo-Carcedo](#) ORCID: [orcid.org/0000-0003-3571-2449](https://orcid.org/0000-0003-3571-2449)<sup>1</sup>,
- [Guillaume Diss](#) ORCID: [orcid.org/0000-0001-9153-4523](https://orcid.org/0000-0001-9153-4523)<sup>1,5</sup> &
- [Ben Lehner](#) ORCID: [orcid.org/0000-0002-8817-1124](https://orcid.org/0000-0002-8817-1124)<sup>1,2,3</sup>

*Nature* volume **604**, pages 175–183 (2022)

- 7254 Accesses
- 242 Altmetric
- [Metrics details](#)

## Subjects

- [Biological techniques](#)
- [Genetics](#)
- [Genomics](#)
- [Molecular evolution](#)
- [Systems biology](#)

## Abstract

Allosteric communication between distant sites in proteins is central to biological regulation but still poorly characterized, limiting understanding, engineering and drug development<sup>1,2,3,4,5,6</sup>. An important reason for this is the lack of methods to comprehensively quantify allostery in diverse proteins. Here we address this shortcoming and present a method that uses deep mutational scanning to globally map allostery. The approach uses an efficient experimental design to infer en masse the causal biophysical effects of mutations by quantifying multiple molecular phenotypes—here we examine binding and protein abundance—in multiple genetic backgrounds and fitting thermodynamic models using neural networks. We apply the approach to two of the most common protein interaction domains found in humans, an SH3 domain and a PDZ domain, to produce comprehensive atlases of allosteric communication. Allosteric mutations are abundant, with a large mutational target space of network-altering ‘edgetic’ variants. Mutations are more likely to be allosteric closer to binding interfaces, at glycine residues and at specific residues connecting to an opposite surface within the PDZ domain. This general approach of quantifying mutational effects for multiple molecular phenotypes and in multiple genetic backgrounds should enable the energetic and allosteric landscapes of many proteins to be rapidly and comprehensively mapped.

This is a preview of subscription content

## Access options

Subscribe to Nature+

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

\$29.99

monthly

[Subscribe](#)

Subscribe to Journal

Get full journal access for 1 year

\$199.00

only \$3.90 per issue

[Subscribe](#)

All prices are NET prices.

VAT will be added later in the checkout.

Tax calculation will be finalised during checkout.

Buy article

Get time limited or full article access on ReadCube.

\$32.00

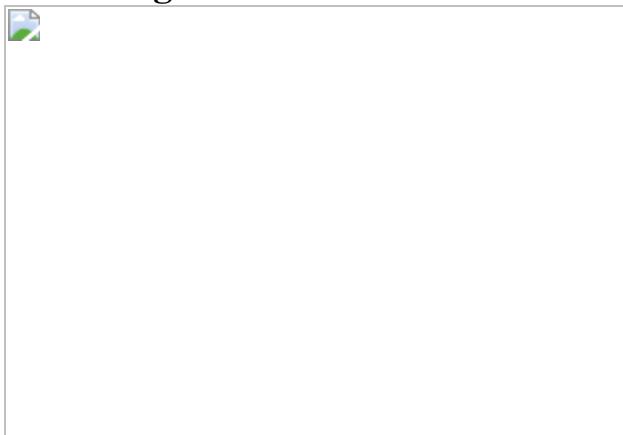
[Buy](#)

All prices are NET prices.

### **Additional access options:**

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

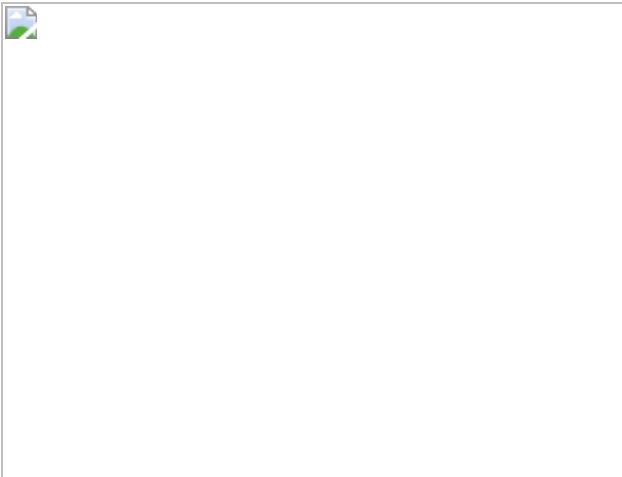
**Fig. 1: ddPCA quantifies the effects of mutations on protein abundance and binding.**



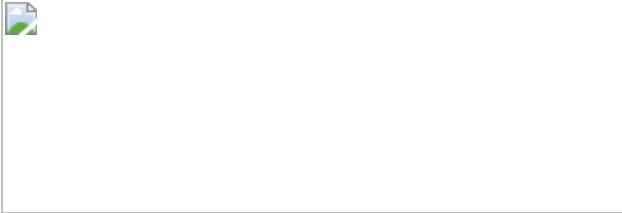
**Fig. 2: From molecular phenotypes to free energy changes.**



**Fig. 3: Binding and folding free energy landscapes of the SH3 and PDZ domains.**



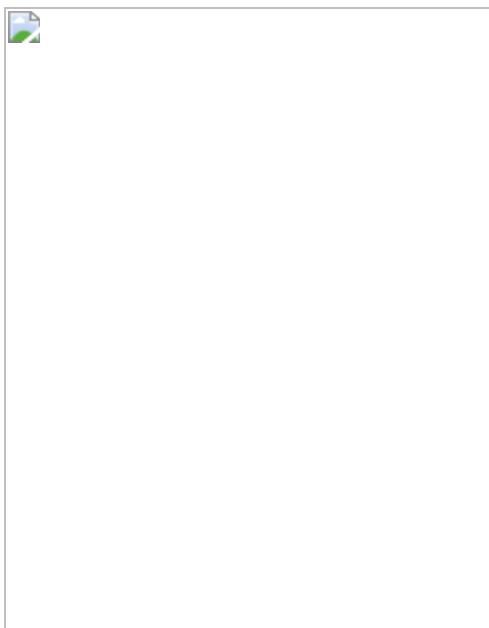
**Fig. 4: Mutational effects on protein stability.**



**Fig. 5: Major allosteric sites in protein binding domains.**



**Fig. 6: Protein surfaces are frequent sites of binding affinity modulation.**



## Data availability

All DNA sequencing data have been deposited in the Gene Expression Omnibus with accession number [GSE184042](#). Protein structures were obtained from the Protein Data Bank with the following accessions: GRB2-SH3, [2VWF](#); PSD95-PDZ3, [1BE9](#); GB1: [1FCC](#); GRB2 homodimer: [1GRI](#),

and the AlphaFold prediction for PSD95 was obtained from the AlphaFold Protein Structure Database with accession [P78352](#).

## Code availability

Source code used to fit thermodynamic models, perform all downstream analyses and to reproduce all figures in this work is available at <https://github.com/lehner-lab/doubledeepms>.

## References

1. Guarnera, E. & Berezovsky, I. N. Allosteric drugs and mutations: chances, challenges, and necessity. *Curr. Opin. Struct. Biol.* **62**, 149–157 (2020).
2. Arkin, M. R., Tang, Y. & Wells, J. A. Small-molecule inhibitors of protein-protein interactions: progressing toward the reality. *Chem. Biol.* **21**, 1102–1114 (2014).
3. Motlagh, H. N., Wrabl, J. O., Li, J. & Hilser, V. J. The ensemble nature of allostery. *Nature* **508**, 331–339 (2014).
4. Xie, J. & Lai, L. Protein topology and allostery. *Curr. Opin. Struct. Biol.* **62**, 158–165 (2020).
5. Kuriyan, J. & Eisenberg, D. The origin of protein interactions and allostery in colocalization. *Nature* **450**, 983–990 (2007).
6. Nussinov, R. & Tsai, C.-J. Allostery in disease and in drug discovery. *Cell* **153**, 293–305 (2013).
7. Monod, J., Changeux, J. P. & Jacob, F. Allosteric proteins and cellular control systems. *J. Mol. Biol.* **6**, 306–329 (1963).
8. Ullmann, A. In memoriam: Jacques Monod (1910–1976). *Genome Biol. Evol.* **3**, 1025–1033 (2011).

9. Halabi, N., Rivoire, O., Leibler, S. & Ranganathan, R. Protein sectors: evolutionary units of three-dimensional structure. *Cell* **138**, 774–786 (2009).
10. Dionne, U. et al. Protein context shapes the specificity of SH3 domain-mediated interactions in vivo. *Nat. Commun.* **12**, 1597 (2021).
11. McCormick, J. W., Russo, M. A., Thompson, S., Blevins, A. & Reynolds, K. A. Structurally distributed surface sites tune allosteric regulation. *eLife* **10**, e68346 (2021).
12. Bandaru, P. et al. Deconstruction of the Ras switching cycle through saturation mutagenesis. *eLife* **6**, e27810 (2017).
13. Reynolds, K. A., McLaughlin, R. N. & Ranganathan, R. Hot spots for allosteric regulation on protein surfaces. *Cell* **147**, 1564–1575 (2011).
14. Oakes, B. L. et al. Profiling of engineering hotspots identifies an allosteric CRISPR–Cas9 switch. *Nat. Biotechnol.* **34**, 646–651 (2016).
15. Leander, M., Yuan, Y., Meger, A., Cui, Q. & Raman, S. Functional plasticity and evolutionary adaptation of allosteric regulation. *Proc. Natl Acad. Sci. USA* **117**, 25445–25454 (2020).
16. Tack, D. S. et al. The genotype-phenotype landscape of an allosteric protein. *Mol. Syst. Biol.* **17**, e10179 (2021).
17. Coyote-Maestas, W., He, Y., Myers, C. L. & Schmidt, D. Domain insertion permissibility-guided engineering of allostery in ion channels. *Nat. Commun.* **10**, 290 (2019).
18. Li, X. & Lehner, B. Biophysical ambiguities prevent accurate genetic prediction. *Nat. Commun.* **11**, 4923 (2020).
19. Otwinowski, J. Biophysical inference of epistasis and the effects of mutations on protein stability and function. *Mol. Biol. Evol.* **35**, 2345–2354 (2018).

20. Fowler, D. M. & Fields, S. Deep mutational scanning: a new style of protein science. *Nat. Methods* **11**, 801–807 (2014).
21. Woodsmith, J. et al. Protein interaction perturbation profiling at amino-acid resolution. *Nat. Methods* **14**, 1213–1221 (2017).
22. Cagiada, M. et al. Understanding the origins of loss of protein function by analyzing the effects of thousands of variants on activity and abundance. *Mol. Biol. Evol.* **38**, 3235–3246 (2021).
23. Domingo, J., Baeza-Centurion, P. & Lehner, B. The causes and consequences of genetic interactions (epistasis). *Annu. Rev. Genom. Hum. Genet.* **20**, 433–460 (2019).
24. Diss, G. & Lehner, B. The genetic landscape of a physical interaction. *eLife* **7**, e32472 (2018).
25. Levy, E. D., Kowarzyk, J. & Michnick, S. W. High-resolution mapping of protein concentration reveals principles of proteome architecture and adaptation. *Cell Rep.* **7**, 1333–1340 (2014).
26. Pelletier, J. N., Arndt, K. M., Plückthun, A. & Michnick, S. W. An in vivo library-versus-library selection of optimized protein-protein interactions. *Nat. Biotechnol.* **17**, 683–690 (1999).
27. Campbell-Valois, F.-X., Tarassov, K. & Michnick, S. W. Massive sequence perturbation of a small protein. *Proc. Natl Acad. Sci. USA.* **102**, 14988–14993 (2005).
28. Tokuriki, N. & Tawfik, D. S. Stability effects of mutations and protein evolvability. *Curr. Opin. Struct. Biol.* **19**, 596–604 (2009).
29. Wei, X. et al. A massively parallel pipeline to clone DNA variants and examine molecular phenotypes of human disease mutations. *PLoS Genet.* **10**, e1004819 (2014).
30. Horovitz, A., Fleisher, R. C. & Mondal, T. Double-mutant cycles: new directions and applications. *Curr. Opin. Struct. Biol.* **58**, 10–17 (2019).

31. Calosci, N. et al. Comparison of successive transition states for folding reveals alternative early folding pathways of two homologous proteins. *Proc. Natl Acad. Sci. USA* **105**, 19241–19246 (2008).
32. Kellogg, E. H., Leaver-Fay, A. & Baker, D. Role of conformational sampling in computing mutation-induced changes in protein structure and stability. *Proteins* **79**, 830–838 (2011).
33. Nisthal, A., Wang, C. Y., Ary, M. L. & Mayo, S. L. Protein stability engineering insights revealed by domain-wide comprehensive mutagenesis. *Proc. Natl Acad. Sci. USA* **116**, 16367–16377 (2019).
34. Laursen, L., Kliche, J., Gianni, S. & Jemth, P. Supertertiary protein structure affects an allosteric network. *Proc. Natl Acad. Sci. USA* **117**, 24294–24304 (2020).
35. Olson, C. A., Wu, N. C. & Sun, R. A comprehensive biophysical description of pairwise epistasis throughout an entire protein domain. *Curr. Biol.* **24**, 2643–2651 (2014).
36. Shoichet, B. K., Baase, W. A., Kuroki, R. & Matthews, B. W. A relationship between protein stability and protein function. *Proc. Natl Acad. Sci. USA* **92**, 452–456 (1995).
37. Redler, R. L., Das, J., Diaz, J. R. & Dokholyan, N. V. Protein destabilization as a common factor in diverse inherited disorders. *J. Mol. Evol.* **82**, 11–16 (2016).
38. Mosca, R., Céol, A. & Aloy, P. Interactome3D: adding structural details to protein networks. *Nat. Methods* **10**, 47–53 (2013).
39. McLaughlin, R. N. Jr et al. The spatial architecture of protein function and adaptation. *Nature* **491**, 138–142 (2012).
40. Wang, J. et al. Mapping allosteric communications within individual proteins. *Nat. Commun.* **11**, 3862 (2020).

41. Zhong, Q. et al. Edgetic perturbation models of human inherited disorders. *Mol. Syst. Biol.* **5**, 321 (2009).
42. Sahni, N. et al. Widespread macromolecular interaction perturbations in human genetic disorders. *Cell* **161**, 647–660 (2015).
43. Kinney, J. B., Murugan, A., Callan, C. G. Jr & Cox, E. C. Using deep sequencing to characterize the biophysical mechanism of a transcriptional regulatory sequence. *Proc. Natl Acad. Sci. USA* **107**, 9158–9163 (2010).
44. Forcier, T. L. et al. Measuring *cis*-regulatory energetics in living cells using allelic manifolds. *eLife* **7**, e40618 (2018).
45. Tareen, A. et al. MAVE-NN: learning genotype–phenotype maps from multiplex assays of variant effect. Preprint at *bioArxiv* <https://doi.org/10.1101/2020.07.14.201475> (2020).
46. Adams, R. M., Mora, T., Walczak, A. M. & Kinney, J. B. Measuring the sequence-affinity landscape of antibodies with massively parallel titration curves. *eLife* **5**, e23156 (2016).
47. Kinney, J. B. & McCandlish, D. M. Massively parallel assays and quantitative sequence–function relationships. *Annu. Rev. Genomics Hum. Genet.* **20**, 99–127 (2019).
48. Skoulidis, F. et al. Sotorasib for lung cancers with *KRAS* p.G12C mutation. *N. Engl. J. Med.* **384**, 2371–2381 (2021).
49. Jumper, J. et al. Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).

## Acknowledgements

This work was funded by European Research Council (ERC) Advanced (883742) and Consolidator (616434) grants, the Spanish Ministry of Science and Innovation (PID2020-118723GB-I00, BFU2017-89488-P, EMBL Partnership, Severo Ochoa Centre of Excellence), the Bettencourt

Schueller Foundation, the AXA Research Fund, Agencia de Gestió d'Ajuts Universitaris i de Recerca (AGAUR, 2017 SGR 1322), and the CERCA Program/Generalitat de Catalunya. J.M.S. was supported by an EMBO Long-Term Fellowship (ALTF 857-2016) and a Marie Skłodowska-Curie Fellowship (752809, EU Commission Horizon 2020). We thank M. Dias, J. Frazer and D. Marks for providing EVE and EVmutation predictions, J. Taipale for motivation and discussion, and all members of the Lehner laboratory for helpful discussions and suggestions, especially P. Baeza-Centurion, X. Li and A. M. New.

## Author information

### Author notes

1. These authors contributed equally: Andre J. Faure, Júlia Domingo, Jörn M. Schmiedel

### Affiliations

1. Center for Genomic Regulation (CRG), The Barcelona Institute of Science and Technology, Barcelona, Spain

Andre J. Faure, Júlia Domingo, Jörn M. Schmiedel, Cristina Hidalgo-Carcedo, Guillaume Diss & Ben Lehner

2. Universitat Pompeu Fabra (UPF), Barcelona, Spain

Ben Lehner

3. Institució Catalana de Recerca i Estudis Avançats (ICREA), Barcelona, Spain

Ben Lehner

4. New York Genome Center (NYGC), New York, NY, USA

Júlia Domingo

5. Friedrich Miescher Institute for Biomedical Research (FMI), Basel,  
Switzerland

Guillaume Diss

## Contributions

J.D., J.M.S., G.D. and B.L. conceived the project and designed the experiments. J.D., J.M.S. and C.H.-C. constructed the mutant libraries. J.D. performed the yeast competition experiments with help from C.H.-C. J.D. constructed the sequencing libraries for next-generation sequencing. A.J.F. led the data analysis with help from J.D. and J.M.S. A.J.F., J.M.S. and B.L. formulated the thermodynamic model. A.J.F. wrote the code to implement and fit the model. B.L., A.J.F. and J.D. wrote the manuscript with input from all authors.

## Corresponding author

Correspondence to [Ben Lehner](#).

## Ethics declarations

## Competing interests

The authors declare no competing interests.

## Peer review

## Peer review information

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

## Additional information

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## Extended data figures and tables

### Extended Data Fig. 1 Performance of thermodynamic models.

**a**, Distribution of the number of double aa substitutions comprising the same single aa substitution in the AbundancePCA (blue) or BindingPCA (red) assays for the GRB2-SH3 (left) and PSD95-PDZ3 (right) protein domains. Median indicated with a dashed line and text label. **b–d**, 2d density plots comparing the ddPCA observed fitness and the model predicted fitness of single (left panels) and double aa substitutions (right panels) for the binding (top panels) and when existing, folding assays (bottom panels) of the GB1 (**b**), GRB2-SH3 (**c**) and PSD95-PDZ3 (**d**) domains.  $R^2$  = proportion variance explained. **e–g**, Same as (**b–d**) but using validation data comprising 10% of double mutants held out during model fitting.

### Extended Data Fig. 2 Performance of thermodynamic models after restricting data to a single phenotype or a single genetic background.

**a**, 2d density plots comparing the observed and predicted fitness of the binding (top panels) and abundance (bottom panels) assays when only the BindingPCA data is used for training the model for GRB2-SH3 (left panels) and PSD95-PDZ3 (right panels). **b**, Same as in (**a**), but only using single mutant data from both binding and abundance assays to fit the models.  $R^2$  = proportion variance explained. **c, d**, Comparisons of inferred free energy changes to previously reported PSD95-PDZ3 mutant *in vitro* measurements where only BindingPCA data (**c**) or single mutants (**d**) were used to fit thermodynamic models. Free energies are from a single model; error bars indicate 95% CI from a Monte Carlo simulation approach ( $n = 10$  experiments) and the regression error bands indicate 95% CI for predictions from a linear model (panel c top:  $n = 22$ , bottom:  $n = 25$ , panel d top:  $n = 32$ , bottom:  $n = 29$ ).  $r$  = Pearson correlation coefficient.

### Extended Data Fig. 3 Performance of thermodynamic models after downsampling and comparisons of inferred free energy changes to smaller-scale datasets of in vitro measurements.

**a**, Dashed lines indicate the relationship between the percentage of fitness variance explained by model predictions with respect to held out validation data (10% of doubles) and the percentage of randomly retained double aa mutants used to train the model in the abundance (blue) or binding (red) assay. Results are shown separately for all protein domains. Solid lines indicate the relationship between the percentage variance explained by inferred free energies with respect to previously reported in vitro measurements for GB1 (Nisthal et al. 2019<sup>33</sup>) and PSD95-PDZ3 (Laursen et al. 2020<sup>34</sup> for  $\Delta\Delta G$  binding, red; Calosci et al. 2008<sup>31</sup> for  $\Delta\Delta G$  folding, blue), where models were trained using varying fractions of randomly downsampled double mutants (x-axis). The top scale indicates the median number of double aa mutants per single aa mutant in the full dataset. **b**, Comparisons of the model-inferred free energy changes to previously reported in vitro measurements for GRB2-SH3 (Malagrinò et al. 2019<sup>56</sup> for  $\Delta\Delta G$  binding and Troilo et al. 2018<sup>57</sup> for  $\Delta\Delta G$  folding) and PSD95-PDZ3 (Chi et al. 2008<sup>58</sup>). Note the modest effect sizes of variants assayed in Malagrinò et al. 2019. Free energies are from a single model; error bars indicate 95%CI from a Monte Carlo simulation approach ( $n = 10$  experiments, in vitro error measurement not provided) and the regression error bands indicate 95% CI for predictions from a linear model (top left:  $n = 11$ , bottom left:  $n = 15$ , top right:  $n = 11$ , bottom right:  $n = 12$ ).  $r$  = Pearson correlation coefficient.

### Extended Data Fig. 4 Correlation of folding free energy changes with computational predictions of mutational effects.

**a**, High confidence inferred folding free energy changes versus corresponding FoldX<sup>59</sup> predictions upon mutation (“PositionScan” command), excluding substitutions involving potentially large increases in mass/volume (at wild-type Glycine, Alanine, Valine) or the replacement of Histidine (whose charge depends on the pH and local chemical environment). **b**, High confidence inferred folding free energy changes

versus corresponding PolyPhen2<sup>60</sup> predictions for amino acid substitutions reachable by single nucleotide substitutions (SNPs). **c**, High confidence inferred folding free energy changes versus corresponding EVE pathogenicity scores<sup>61</sup>. **d**, Same as in (**c**), but scores are based on evolutionary couplings<sup>62</sup>.  $r$  = Pearson correlation coefficient.

**Extended Data Fig. 5 Binding and folding free energy landscapes of the GB1 domain and biophysical mechanism of mutations that affect binding.**

**a, b**, Heatmaps showing inferred changes in free energies of binding (**a**) and folding (**b**) for the GB1 domain. The final row in each heatmap indicates the minimal distance to the ligand (considering the side chain heavy atoms or the alpha carbon atoms in the case of glycine). Free energy changes of ligand-proximal residues (ligand distance  $< 5 \text{ \AA}$ ) are boxed. Low confidence estimates are indicated with dots (95% CI  $\geq 1 \text{ kcal/mol}$ ). Free energy changes more extreme than  $\pm 2.5$  were set to this limit. **c**, Scatter plot comparing binding and folding free energy changes of mutations in the core, surface and binding interface. Contours indicate estimates of 2D densities with 6 contour bins. **d**, Distribution of binding (red) and folding (blue) free energy changes. **e**, Percentage of mutations that significantly decrease (top) or increase (bottom) fitness in the binding assay (FDR = 0.05) categorised by their biophysical mechanism. Pleiotropic mutations have significant changes in free energies of both folding and binding (FDR = 0.05) and are classified as either synergistic or antagonistic depending on whether their effects are in the same or different direction respectively. **f**, Changes in free energy of binding (blue) or folding (red) of single aa substitutions with different fitness effects in the binding assay for the three protein domains. **g**, Percentage of core, surface or ligand binding mutations that significantly decrease (top) or increase (bottom) fitness in the binding assay (FDR = 0.05) categorised by their biophysical mechanism. Pleiotropic mutations have significant changes in free energies of both folding and binding (FDR = 0.05) and are classified as either synergistic or antagonistic depending on whether their effects are in the same or different direction respectively.

## Extended Data Fig. 6 GB1 mutational effects on protein stability and characterisation of surface de-stabilizing residues.

**a**, 3D structure of GB1 (PDB entry 1FCC) where residue atoms are coloured by the position-wise average change in the free energy of folding. The FC domain of the human Immunoglobulin G is shown as black sticks. **b**, Violin plots indicating distributions of confident changes in free energy of folding ( $n = 898$ ; \*\*\* $P < 2.2e-16$ , two-sided Mann-Whitney U test comparing mutations in the core versus the remainder). **c**, Anti-correlation between the position-wise average change in free energy of folding and the solvent exposure of the corresponding residue (RSASA) in GB1. Error bars indicate 95% CI ( $n = 19$ ).  $r$  = Pearson correlation coefficient. **d**, Percentage of core, surface or binding-interface residues in GB1 shown separately for de-stabilizing residues (positions with  $\geq 5$  stabilizing mutations, folding  $\Delta\Delta G < 0$ , FDR = 0.05) and the remainder. Inset numbers are total counts. **e**, Violin plots indicating evolutionary conservation scores (from a multiple sequence alignment of 185, 8,852, 276,481 homologous sequences of the GB1, GRB2-SH3 and PSD95-PDZ3 domains, respectively) shown separately for surface de-stabilizing residues and remaining surface or core residues. **f**, Violin plots indicating hydrophobicity score distributions shown separately for surface de-stabilizing residues and remaining surface or core residues. **g**, 3D structures of the GRB2-SH3 and PSD95-PDZ3 domains (grey cartoons) with the side-chains of surface de-stabilizing residues highlighted in green sticks. Ligands are shown as black sticks. In the insets, in yellow is shown the SH2 domains of the second monomer of GRB2 when found in dimeric form (left, PDB entry 1GRI)<sup>63</sup>, and relevant proximal portions of PSD95 C-terminal to the PDZ3 domain (middle and right, PDB entry 1BE9 and AlphaFold Protein Structure Database entry P78352).

## Extended Data Fig. 7 Major allosteric sites in the GB1 domain and changes in free energy of binding in ligand binding interfaces.

**a**, 3D structures of the protein G B1 domain where residue atoms are coloured by the position-wise average absolute change in the free energy of

binding. The FC domain of the human Immunoglobulin G is shown as black sticks. **b**, GB1 domain structure with binding-interface residues (ligand distance < 5 Å) highlighted in red and major allosteric site residues highlighted in orange **c**, Relationship between the position-wise average absolute change in free energy of binding and the distance to the ligand (minimal side chain heavy atom distance) in the GB1 domain. Major allosteric sites (yellow) are defined as non-binding-interface residues with weighted average absolute change in free energy of binding higher than the average of binding-interface residue mutations (red). **d**, ROC curves for predicting ligand-contacting residues (ligand distance < 5 Å) using (weighted) mean absolute binding  $\Delta\Delta G$  considering all variants or those with confident inferred free energies (conf.). AUC = Area Under the Curve. **e**, Inferring changes in free energy of binding provides insights into the interactions that mediate binding between GRB2-SH3 and GAB2 peptide, and how mutations disrupt binding. F7 and Y51 of the GRB2-SH3 domain contact P3 and P4 of the GAB2 peptide through aromatic-proline interactions (left heatmap). In these two positions, only mutations to Y, F, Q and H, which can interact with proline through aromatic-proline or amino-aromatic interactions, are tolerated, while all other amino acid substitutions result in decreased binding affinity (positive binding  $\Delta\Delta G$ ). Residue M46 can tolerate all amino acid substitutions except to positively charged residues (right heatmap). The closest residue of GAB2 is a lysine, and so a repulsive electrostatic interaction likely occurs when a positively charged amino acid occupies position 46 of the SH3 domain (binding  $\Delta\Delta G$  of 2.1 and 1.99 for M46K and M46R respectively). **f**, ROC curves for predicting ligand contacting residues using (weighted) mean BindingPCA or AbundancePCA fitness.

### [Extended Data Fig. 8 Changes in fitness and free energy of binding and folding of major allosteric sites and allosteric mutations.](#)

**a**, Scatter plots of single aa substitutions' changes in free energy of binding and folding for the GB1 (left panel), GRB2-SH3 (middle panel) and PSD95-PDZ3 (right panel) protein domains. Variants are coloured by aa position if found in a major allosteric site. Free energies are from a single model; error bars indicate 95% CI from a Monte Carlo simulation approach

(n = 10 experiments). **b**, Scatter plots comparing abundance and binding fitness of single aa substitutions in GRB2-SH3 (left panel) and PSD95-PDZ3 (right panel). Variants are coloured by aa position if found in a major allosteric site. Data are presented as mean values and error bars indicate 95% CI (n = 3 biological replicates). The red line indicates the model-derived relationship between abundance and binding fitness in the absence of a change in the free energy of binding. **c**, Scatter plots of single aa substitutions' changes in free energy of binding and folding for GB1 (left panel), GRB2-SH3 (middle panel) and PSD95-PDZ3 (right panel). Variants are coloured by aa position if found in a major allosteric site (yellow) or in a position that has allosteric mutations (green). Free energies are from a single model; error bars indicate 95% CI from a Monte Carlo simulation approach (n = 10 experiments). **d**, Scatter plots comparing abundance and binding fitness of single aa substitutions in GRB2-SH3 (left panel) and PSD95-PDZ3 (right panel). Variants are coloured by aa position if found in a major allosteric site (yellow) or in a position that has allosteric mutations (green). Data are presented as mean values and error bars indicate 95% CI (n = 3 biological replicates). The red line indicates the model-derived relationship between abundance and binding fitness in the absence of a change in the free energy of binding.

### **Extended Data Fig. 9 Allosteric mutations in GB1 and enrichment of allosteric mutations in literature allosteric networks and specific residue types and classes.**

**a**, Domain structure of GB1 with surface allosteric sites and surface residues with allosteric mutations highlighted in orange and green respectively. The FC domain of the human Immunoglobulin G is shown as black sticks. **b**, Scatter plot showing the binding free energy changes of all mutations and coloured according to residue position: allosteric site (orange), orthosteric site/mutation (red), core allosteric mutation (blue), surface allosteric mutation (green). **c**, Percentage of allosteric mutations per residue versus ligand proximity, excluding sites within the binding interface. Points are coloured according to residue position and major allosteric sites are indicated (see legend).  $\rho$  = Spearman rank correlation coefficient. **d**, Total numbers of mutations decreasing or increasing binding fitness (i.e. the fraction of bound protein complex) beyond the indicated

minimum or maximum thresholds (x-axis; two-sided Z-test  $P < 0.05$ ) respectively. **e**, Enrichment of allosteric mutations in sets of residues defined by previously reported allosteric networks in PSD95-PDZ3: McLaughlin et al. 2012<sup>39</sup>, Salinas et al. 2018<sup>64</sup>, Gerek et al. 2011<sup>65</sup>, Kumawat et al. 2017<sup>66</sup>, Gianni et al. 2011<sup>67</sup>, Kalescky et al. 2015<sup>68</sup>, Du et al. 2010<sup>69</sup>, Kaya et al. 2013<sup>70</sup>. The enrichment ( $\log_2$  odds ratio) corresponding to a 2x2 contingency table is shown on the x-axis and the associated  $P$  value from a two-sided Fisher's Exact Test is indicated. Residues within the binding interface (ligand distance  $< 5 \text{ \AA}$ ) were ignored. Original literature allosteric network sizes are shown in parentheses. **f-g**, Same as (e) except sets of residues are defined by the identity of the WT or mutant amino acid (see legend) or their physicochemical properties (hydrophobic i.e. A, V, I, L, M, F, Y, W or charged i.e. R, H, K, D, E). Results are shown for all residues outside the binding interface (**f**) and further restricted to those residues in beta strands or helices i.e. not within loops/turns (**g**). Sets are ranked by their mean effect across the three protein domains.

### **Extended Data Fig. 10 Comparisons to computationally predicted allosteric coupling scores and mutational biases towards increased or decreased binding given the position in the domain structure.**

**a**, Percentage of allosteric mutations per residue versus allosteric coupling scores estimated by a network-based perturbation propagation algorithm<sup>40</sup>, where residues in the binding interface (ligand distance  $< 5 \text{ \AA}$ ) are omitted as they represent the query set. Residues immediately adjacent to binding-interface residues in the linear aa sequence (i.e. backbone-backbone contacts which are disregarded by the Ohm algorithm) were given the maximum allosteric coupling score (1.0). Major allosteric sites (in yellow) and Spearman rank correlation coefficients ( $\rho$ ) are indicated. **b**, Total numbers of mutations decreasing or increasing the free energy of binding beyond the indicated minimum or maximum thresholds (x-axis; two-sided Z-test  $P < 0.05$ ) respectively, stratified by position in the structure considering all variants (regardless of the confidence of inferred free energies).

# **Supplementary information**

## **Supplementary Methods**

This file contains Methods, supplementary text, equations and additional references.

## **Reporting Summary**

## **Peer Review File**

## **Supplementary Table 1**

Primers used in this study.

## **Supplementary Table 2**

Gene blocks used in this study.

## **Supplementary Table 3**

Experimental details and numbers of the mutagenesis libraries in this study.

## **Supplementary Table 4**

Illumina indexed primers combinations used in this study to demultiplex samples after deep sequencing.

## **Supplementary Table 5**

Degenerate NNK oligonucleotides used for the GRB2-SH3 and PSD95-PDZ3 nicking mutagenesis libraries.

## **Supplementary Table 6**

Fitness estimates for GB1, GRB2-SH3 and PSD95-PDZ3.

## Supplementary Table 7

Inferred folding and binding free energy changes and associated annotations for GB1, GRB2-SH3 and PSD95-PDZ3.

## Rights and permissions

### Reprints and Permissions

## About this article

### Cite this article

Faure, A.J., Domingo, J., Schmiedel, J.M. *et al.* Mapping the energetic and allosteric landscapes of protein binding domains. *Nature* **604**, 175–183 (2022). <https://doi.org/10.1038/s41586-022-04586-4>

- Received: 14 September 2021
- Accepted: 25 February 2022
- Published: 06 April 2022
- Issue Date: 07 April 2022
- DOI: <https://doi.org/10.1038/s41586-022-04586-4>

## Share this article

Anyone you share the following link with will be able to read this content:

[Get shareable link](#)

Sorry, a shareable link is not currently available for this article.

[Copy to clipboard](#)

Provided by the Springer Nature SharedIt content-sharing initiative

## **Democratizing the mapping of gene mutations to protein biophysics**

- Debora S. Marks
- Stephen W. Michnick

News & Views 06 Apr 2022

---

This article was downloaded by **calibre** from <https://www.nature.com/articles/s41586-022-04586-4>

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

- Article
- [Published: 03 February 2022](#)

# Structure of the NLRP3 decamer bound to the cytokine release inhibitor CRID3

- [Inga V. Hochheiser](#) ORCID: [orcid.org/0000-0001-6649-2624](https://orcid.org/0000-0001-6649-2624)<sup>1</sup>,
- [Michael Pilsl](#) ORCID: [orcid.org/0000-0002-5993-4735](https://orcid.org/0000-0002-5993-4735)<sup>2</sup>,
- [Gregor Hagelueken](#)<sup>1</sup>,
- [Jonas Moecking](#) ORCID: [orcid.org/0000-0001-5914-1359](https://orcid.org/0000-0001-5914-1359)<sup>1</sup>,
- [Michael Marleaux](#)<sup>1</sup>,
- [Rebecca Brinkschulte](#)<sup>1</sup>,
- [Eicke Latz](#) ORCID: [orcid.org/0000-0003-1488-5666](https://orcid.org/0000-0003-1488-5666)<sup>3</sup>,
- [Christoph Engel](#) ORCID: [orcid.org/0000-0001-8451-3599](https://orcid.org/0000-0001-8451-3599)<sup>2</sup> &
- [Matthias Geyer](#) ORCID: [orcid.org/0000-0002-7718-5002](https://orcid.org/0000-0002-7718-5002)<sup>1</sup>

[Nature](#) volume 604, pages 184–189 (2022)

- 8677 Accesses
- 2 Citations
- 128 Altmetric
- [Metrics details](#)

## Subjects

- [Cryoelectron microscopy](#)
- [Inflammasome](#)

# Abstract

NLRP3 is an intracellular sensor protein that when activated by a broad spectrum of exogenous and endogenous stimuli leads to inflammasome formation and pyroptosis<sup>1,2</sup>. The conformational states of NLRP3 and the way antagonistic small molecules act at the molecular level remain poorly understood<sup>2,3</sup>. Here we report the cryo-electron microscopy structures of full-length human NLRP3 in its native form and complexed with the inhibitor CRID3 (also named MCC950)<sup>4</sup>. Inactive, ADP-bound NLRP3 is a decamer composed of homodimers of intertwined leucine-rich repeat (LRR) domains that assemble back-to-back as pentamers. The NACHT domain is located at the apical axis of this spherical structure. One pyrin domain dimer is in addition formed inside the LRR cage. Molecular contacts between the concave sites of two opposing LRR domains are mediated by an acidic loop that extends from an LRR transition segment. Binding of CRID3 considerably stabilizes the NACHT and LRR domains relative to each other. CRID3 binds into a cleft, connecting four subdomains of the NACHT with the transition LRR. Its central sulfonylurea group interacts with the Walker A motif of the NLRP3 nucleotide-binding domain and is sandwiched between two arginine residues, which explains the specificity of NLRP3 for this chemical entity. With the determination of the binding site of this key therapeutic agent, specific targeting of NLRP3 for the treatment of autoinflammatory and autoimmune diseases and rational drug optimization is within reach.

This is a preview of subscription content

## Access options

Subscribe to Nature+

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

\$29.99

monthly

## Subscribe

Subscribe to Journal

Get full journal access for 1 year

\$199.00

only \$3.90 per issue

## Subscribe

All prices are NET prices.

VAT will be added later in the checkout.

Tax calculation will be finalised during checkout.

Buy article

Get time limited or full article access on ReadCube.

\$32.00

## Buy

All prices are NET prices.

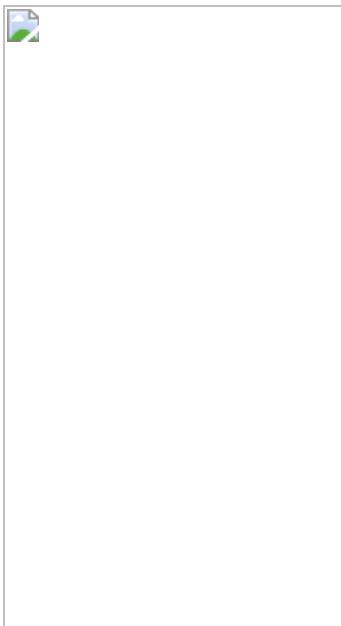
## **Additional access options:**

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

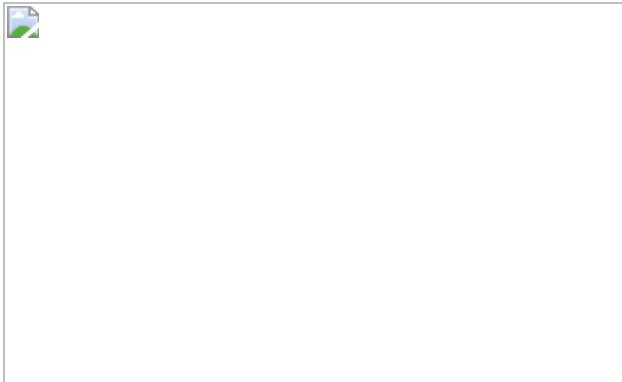
**Fig. 1: Structure of the NLRP3 decamer.**



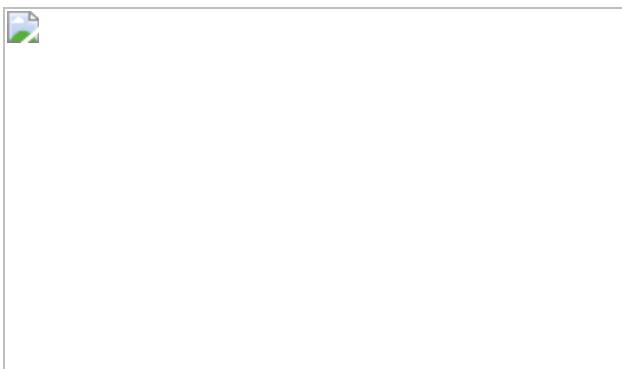
**Fig. 2: The acidic loop interacts with the concave side of the LRR domain.**



**Fig. 3: Interfaces in the NLRP3 decamer assembly.**



**Fig. 4: CRID3 binds into a crevice in NLRP3 interacting with five subdomains.**



## Data availability

The cryo-EM density reconstructions and models have been deposited in the Electron Microscopy Data Bank (EMDB) (accession codes [EMD-13687](#) for the NLRP3 apo decamer and [EMD-13684](#), [EMD-13685](#), [EMD-13686](#), [EMD-13692](#), [EMD-13693](#) and [EMD-13699](#) for the NLRP3–ADP–CRID3 decamer), and in the PDB under accession code [7PZC](#). All data are available in the Article or its [Supplementary Information](#) files. [Source data](#) are provided with this paper.

## References

1. Schroder, K. & Tschopp, J. The inflammasomes. *Cell* **140**, 821–832 (2010).
2. Swanson, K. V., Deng, M. & Ting, J. P. The NLRP3 inflammasome: molecular activation and regulation to therapeutics. *Nat. Rev. Immunol.*

**19**, 477–489 (2019).

3. Latz, E., Xiao, T. S. & Stutz, A. Activation and regulation of the inflammasomes. *Nat. Rev. Immunol.* **13**, 397–411 (2013).
4. Laliberte, R. E. et al. Glutathione S-transferase omega 1-1 is a target of cytokine release inhibitory drugs and may be responsible for their effect on interleukin-1 $\beta$  posttranslational processing. *J. Biol. Chem.* **278**, 16567–16578 (2003).
5. Song, N. et al. NLRP3 phosphorylation is an essential priming event for inflammasome activation. *Mol. Cell* **68**, 185–197 (2017).
6. Schmid-Burgk, J. L. et al. A genome-wide CRISPR (clustered regularly interspaced short palindromic repeats) screen identifies NEK7 as an essential component of NLRP3 inflammasome activation. *J. Biol. Chem.* **291**, 103–109 (2016).
7. Shi, H. et al. NLRP3 activation and mitosis are mutually exclusive events coordinated by NEK7, a new inflammasome component. *Nat. Immunol.* **17**, 250–258 (2016).
8. He, Y., Zeng, M. Y., Yang, D., Motro, B. & Núñez, G. NEK7 is an essential mediator of NLRP3 activation downstream of potassium efflux. *Nature* **530**, 354–357 (2016).
9. Mangan, M. S. J. et al. Targeting the NLRP3 inflammasome in inflammatory diseases. *Nat. Rev. Drug Discov.* **17**, 588–606 (2018).
10. Primiano, M. J. et al. Efficacy and pharmacology of the NLRP3 inflammasome inhibitor CP-456,773 (CRID3) in murine models of dermal and pulmonary inflammation. *J. Immunol.* **197**, 2421–2433 (2016).
11. Harrison, D. et al. Discovery of a series of ester-substituted NLRP3 inflammasome inhibitors. *Bioorg. Med. Chem. Lett.* **30**, 127560 (2020).

12. Coll, R. C. et al. A small-molecule inhibitor of the NLRP3 inflammasome for the treatment of inflammatory diseases. *Nat. Med.* **21**, 248–255 (2015).
13. Vande Walle, L. et al. MCC950/CRID3 potently targets the NACHT domain of wild-type NLRP3 but not disease-associated mutants for inflammasome inhibition. *PLoS Biol.* **17**, e3000354 (2019).
14. Corcoran, S. E., Halai, R. & Cooper, M. A. Pharmacological inhibition of the Nod-like receptor family pyrin domain containing 3 inflammasome with MCC950. *Pharmacol. Rev.* **73**, 968–1000 (2021).
15. Sharif, H. et al. Structural mechanism for NEK7-licensed activation of NLRP3 inflammasome. *Nature* **570**, 338–343 (2019).
16. Kobe, B. & Deisenhofer, J. Mechanism of ribonuclease inhibition by ribonuclease inhibitor protein based on the crystal structure of its complex with ribonuclease A. *J. Mol. Biol.* **264**, 1028–1043 (1996).
17. Bae, J. Y. & Park, H. H. Crystal structure of NALP3 protein pyrin domain (PYD) and its implications in inflammasome assembly. *J. Biol. Chem.* **286**, 39528–39536 (2011).
18. Maekawa, S., Ohto, U., Shibata, T., Miyake, K. & Shimizu, T. Crystal structure of NOD2 and its implications in human disease. *Nat. Commun.* **7**, 11813 (2016).
19. Hu, Z. et al. Crystal structure of NLRC4 reveals its autoinhibition mechanism. *Science* **341**, 172–175 (2013).
20. Tenthorey, J. L. et al. The structural basis of flagellin detection by NAIP5: a strategy to limit pathogen immune evasion. *Science* **358**, 888–893 (2017).
21. Erzberger, J. P. & Berger, J. M. Evolutionary relationships and structural mechanisms of AAA+ proteins. *Annu. Rev. Biophys. Biomol. Struct.* **35**, 93–114 (2006).

22. Hoss, F. et al. Alternative splicing regulates stochastic NLRP3 activity. *Nat. Commun.* **10**, 3238 (2019).
23. Touisou, I. et al. Infevers: an evolving mutation database for auto-inflammatory syndromes. *Hum. Mutat.* **24**, 194–198 (2004).
24. Coll, R. C. et al. MCC950 directly targets the NLRP3 ATP-hydrolysis motif for inflammasome inhibition. *Nat. Chem. Biol.* **15**, 556–559 (2019).
25. Tapia-Abellán, A. et al. MCC950 closes the active conformation of NLRP3 to an inactive state. *Nat. Chem. Biol.* **15**, 560–564 (2019).
26. Wendler, P., Ciniawsky, S., Kock, M. & Kube, S. Structure and function of the AAA+ nucleotide binding pocket. *Biochim. Biophys. Acta* **1823**, 2–14 (2012).
27. Compan, V. et al. Cell volume regulation modulates NLRP3 inflammasome activation. *Immunity* **37**, 487–500 (2012).
28. Zhang, Z. et al. Protein kinase D at the Golgi controls NLRP3 inflammasome activation. *J. Exp. Med.* **214**, 2671–2693 (2017).
29. Green, J. P. et al. Chloride regulates dynamic NLRP3-dependent ASC oligomerization and inflammasome priming. *Proc. Natl Acad. Sci. USA* **115**, E9371–E9380 (2018).
30. Hafner-Bratkovič, I. et al. NLRP3 lacking the leucine-rich repeat domain can be fully activated via the canonical inflammasome pathway. *Nat. Commun.* **9**, 5182 (2018).
31. Hochheiser, I. V. et al. Directionality of PYD filament growth determined by the transition of NLRP3 nucleation seeds to ASC elongation. Preprint at <https://doi.org/10.1101/2021.11.25.470035> (2021).
32. Andreeva, L. et al. Full-length NLRP3 forms oligomeric cages to mediate NLRP3 sensing and activation. Preprint at

<https://doi.org/10.1101/2021.09.12.459968> (2021).

33. Schwaid, A. G. & Spencer, K. B. Strategies for targeting the NLRP3 inflammasome in the clinical and preclinical space. *J. Med. Chem.* **64**, 101–122 (2021).
34. Mastronarde, D. N. Automated electron microscope tomography using robust prediction of specimen movements. *J. Struct. Biol.* **152**, 36–51 (2005).
35. Zivanov, J. et al. New tools for automated high-resolution cryo-EM structure determination in RELION-3. *eLife* **7**, e42166 (2018).
36. Waterhouse, A. et al. SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Res.* **46**, W296–W303 (2018).
37. Pettersen, E. F. et al. UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).
38. Casañal, A., Lohkamp, B. & Emsley, P. Current developments in Coot for macromolecular model building of electron cryo-microscopy and crystallographic data. *Protein Sci.* **29**, 1055–1064 (2020).
39. Dekker, C. et al. Crystal structure of NLRP3 NACHT domain with an inhibitor defines mechanism of inflammasome inhibition. *J. Mol. Biol.* **433**, 167309 (2021).
40. Hochheiser, I. V. et al. Cryo-EM structure of the NLRP3 decamer bound to the cytokine release inhibitory drug CRID3. Preprint at <https://doi.org/10.1101/2021.07.22.453353> (2021).
41. Liebschner, D. et al. Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix. *Acta Crystallogr. D* **75**, 861–877 (2019).
42. Croll, T. I. ISOLDE: a physically realistic environment for model building into low-resolution electron-density maps. *Acta Crystallogr.*

*D* **74**, 519–530 (2018).

43. Prisant, M. G., Williams, C. J., Chen, V. B., Richardson, J. S. & Richardson, D. C. New tools in MolProbity validation: CaBLAM for cryoEM backbone, UnDowser to rethink “waters,” and NGL Viewer to recapture online 3D graphics. *Protein Sci.* **29**, 315–329 (2020).
44. Krissinel, E. & Henrick, K. Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol.* **372**, 774–797 (2007).
45. Keuler, T. et al. Development of fluorescent and biotin probes targeting NLRP3. *Front. Chem.* **9**, 642273 (2021).
46. Kofoed, E. M. & Vance, R. E. Blue native polyacrylamide gel electrophoresis to monitor inflammasome assembly and composition. *Methods Mol. Biol.* **1040**, 169–183 (2013).
47. Sester, D. P. et al. A novel flow cytometric method to assess inflammasome formation. *J. Immunol.* **194**, 455–462 (2015).
48. Zhang, L. et al. Cryo-EM structure of the activated NAIP2-NLRC4 inflammasome reveals nucleated polymerization. *Science* **350**, 404–409 (2015).
49. Park, K. et al. Control of repeat-protein curvature by computational protein design. *Nat. Struct. Mol. Biol.* **22**, 167–174 (2015).
50. Robert, X. & Gouet, P. Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res.* **42**, W320–W324 (2014).

## Acknowledgements

Access to cryo-EM time was supported by iNEXT-Discovery grant PID 11967 for the NLRP3–CRID3 sample and INSTRUCT grant PID 11340 for the NLRP3 apo sample. We thank W. Hagen, F. Weis and A. Durand for cryo-EM data recordings; H. Stark for initial advice on electron microscopy sample preparation; and A. Steiner for instructions on cell assays. C.E.

thanks B. Böttcher for facility access, R. Witzgall for JEM2100F access, G. Meister for shared infrastructure, G. Lehmann and N. Eichner for IT support and all members of the Structural Biochemistry Group of the Regensburg Center for Biochemistry for help and support. C.E. and M.P. acknowledge funding by the Emmy Noether Programme (DFG grant EN 1204/1-1 to C.E.) and SFB 960 (TP-A8 to C.E.). M.G. and E.L. are funded by the Deutsche Forschungsgemeinschaft under Germany's Excellence Strategy—EXC2151–390873048.

## Author information

### Affiliations

1. Institute of Structural Biology, University of Bonn, Bonn, Germany

Inga V. Hochheiser, Gregor Hagelueken, Jonas Moecking, Michael Marleaux, Rebecca Brinkschulte & Matthias Geyer

2. Structural Biochemistry Group, Regensburg Center for Biochemistry, University of Regensburg, Regensburg, Germany

Michael Pilsl & Christoph Engel

3. Institute of Innate Immunity, University of Bonn, Bonn, Germany

Eicke Latz

### Contributions

I.V.H. performed all experiments and data analyses, except for the following. J.M. characterized the interface mutants in cells. M.M. performed SPR experiments. R.B. identified peak 2. I.V.H. and M.P. prepared cryo-EM samples. I.V.H., M.P., G.H. and C.E. processed cryo-EM data. G.H., M.G., M.P. and C.E. built the model. I.V.H., M.P., G.H. C.E. and M.G. interpreted the data. E.L. provided reagents for cell-based assays, and contributed to discussions. C.E. supervised cryo-EM sample preparation, optimization and data interpretation and contributed to funding. M.G.

designed and supervised research and wrote the manuscript. All authors discussed and commented on the manuscript.

## Corresponding author

Correspondence to [Matthias Geyer](#).

## Ethics declarations

### Competing interests

M.G. and E.L. are co-founders and consultants of IFM Therapeutics. The other authors declare no competing interests.

## Peer review

### Peer review information

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

## Additional information

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## Extended data figures and tables

### [Extended Data Fig. 1 Preparation of full length, wild-type human NLRP3 for electron microscopy analyses.](#)

**a**, Analytical SEC of recombinant, human MBP–NLRP3 (3–1036) reveals two elution peaks, one close to the void volume (peak 1) and one at a size larger than 1 MDa (peak 2). The peak ratio is shifted from peak 1 under

standard buffer conditions (50 mM HEPES pH 7.5, 150 mM NaCl, 10 mM MgCl<sub>2</sub>, 1 mM ADP, 0.5 mM TCEP) towards peak 2 by addition of CRID3 (10 μM CRID3). Analytical SEC runs were performed on a Superose 6 Increase 10/300 GL column with 700 μl applied at a concentration of 3 mg/ml (repeated >3 times). **b**, Coomassie-stained SDS PAGE analysis of purified MBP-NLRP3 (repeated >3 times). **c**, SEC–MALS experiment of MBP-NLRP3 Peak 2 plus ADP on a Superose 6 Increase 10/300 GL column. The calculated mass of an MBP-NLRP3 monomer is 158 kDa. **d**, SEC–MALS experiment of MBP-NLRP3 Peak 2 plus ADP plus CRID3 on a Superose 6 Increase 10/300 GL column. The average apparent mass of Peak 2 from both experiments equals 10-times the calculated mass of the protein. **e**, Analytical SEC of recombinant, human NLRP3 (3–1036) after TEV digestion using a Superose 6 Increase 3.2/300 column. **f**, Negative-stain micrograph of TEV-digested, gel filtered NLRP3 plus CRID3 plus BS3 Peak 2 elution fraction that was used for subsequent cryo-EM grid preparation (right, one among a few hundred images). An SDS-PAGE analysis of the protein w/o BS3 is shown left. The scale-bar equals 200 nm. **g**, Exposure of NLRP3 peak 2 to ATP does not convert the decamer assembly into another state. MBP-NLRP3 expressed in *Sf9* cells and purified by affinity chromatography in the absence of any nucleotide elutes in two peaks (top). Incubation of MBP-NLRP3 peak 2 (from top) with 1 mM ATP•Mg<sup>2+</sup> for 2 h at 4 °C does not lead to a change in the elution profile, indicating the autoinhibition of the decamer assembly (bottom). **h**, THP-1 cells contain an oligomeric endogenous NLRP3 species. Immunoblotting of BN-PAGE of THP-1 cell lysate supernatant and pellet fractions from PMA-differentiated, untreated and LPS-treated wild-type cells pre-incubated with or without CRID3 (repeated 2 times). The supernatant of LPS-treated and BS3 cross-linked cell lysates contains an oligomeric NLRP3 fraction in the presence or absence of CRID3 (left). In the pellet fractions of LPS-treated cell lysates only an oligomeric NLRP3 species was detected at a molecular mass around 1.15 MDa, in line with the formation of a decamer (middle). Anti-NLRP3 Cryo-2 and anti-β actin antibodies were used for the WB analysis. BN-PAGE analysis of recombinant human (MBP-)NLRP3 protein expressed in *Sf9* cells and purified by affinity chromatography and SEC (right). Source data are provided as a source data file.

[Source data](#)

**Extended Data Fig. 2 Biochemical characterization of NLRP3 peak 1 and peak 2.**

**a**, Analytical SEC of NLRP3 (3–1036) mutants S198E and S198A. The phospho-mimetic S198E mutant shows a decreased content of peak 2. **b**, Using a multi-cycle turnover reverse-phase HPLC assay with 3 μM protein and 100 μM ATP concentration at 25 °C, NLRP3 peak 1 shows an ~10-times higher ATP turnover number than peak 2. **c**, Negative stain electron micrographs of TEV-cleaved human NLRP3 peak 1 or peak 2, co-incubated with ASC-mCherry in a 1:10 molar ratio or ASC-mCherry incubated alone for 3 min at 25 °C. All samples show a comparable prevalence of ASC filament formation events. Images are representative of two independent experiments. The scale bars are 500 nm. **d**, Pull-down experiments of *Sf9* cell expressed GST-NEK7 (bait) with MBP-NLRP3 peak 2 (prey) after incubation for 2 h at 4 °C. Experiments were repeated 2 times. **e**, *Sf9* cell expressed GST-NEK7 and MBP-NLRP3 (peak 2) were separately purified and co-incubated (2x molar excess of GST-NEK7) at 4 °C for 4 h and subjected to gel filtration on a Superose 6 increase 10/300 GL column (repeated 3 times). Both recombinant proteins elute separately from each other. **f**, Elution profile (S6 increase 10/300 GL column) and SDS-PAGE analysis of *Sf9* cell co-expressed MBP-NLRP3 and GST-NEK7 (repeated 2 times). An elution peak containing both proteins indicates the formation of a NLRP3–NEK7 heterodimer complex. Source data are provided as a source data file.

[Source data](#)

**Extended Data Fig. 3 Cryo-EM data processing of apo-NLRP3 and the NLRP3–CRID3 complex.**

**a**, Representative cryo-EM micrograph of the human NLRP3 oligomer in the apo state eluting as Peak 2 (replicated more than 35,000 times). Scale bar, 50 nm. **b**, Representative 2D class averages. **c**, Processing tree describing particle classification of the NLRP3 cryo-EM data in RELION3.1<sup>35</sup>. **d**, Top, tilted, and side views of the cryo-EM density of the

NLRP3 protein w/o CRID3 at ~10 Å resolution. The 3.9 Å structure of the NLRP3–CRID3 complex is fitted into the density map showing that the apo and the inhibitor-bound NLRP3 decamer structures do not exhibit large conformational rearrangements. **e**, Representative 2D class averages of the NLRP3–CRID3 complex. **f**, Processing tree of cryo-EM data in cryo-SPARC leading to 3 ab initio models. **g**, The second model, which was based on 409,755 particles, was further refined with D5 symmetry resulting in a 3.8 Å resolution map. **h**, Model 2 was further processed into two 3D classes, one of which contained 237,390 particles. Non-uniform (NU) refinement of the latter class in C1 symmetry led to a 4.8 Å resolution map. The centre region of the NLRP3 cage showed extra density that was fitted with a PYD dimer as described in the main text.

### Extended Data Fig. 4 Cryo-EM data processing of the NLRP3–CRID3 complex.

**a**, Representative cryo-EM micrograph of the human NLRP3 decamer in complex with CRID3 (replicated more than 20,000 times). Scale bar, 50 nm. **b**, **c**, Representative 2D class averages. **d**, **e**, Processing tree of cryo-EM data in RELION3.1<sup>35</sup> as described in the Methods section and gold standard FSC plots of masked final maps (grey) as calculated from phenix-mtriage<sup>40</sup>. **f–h**, Orientation distribution of particles and heat maps of local resolution estimations.

### Extended Data Fig. 5 Quality of cryo-EM densities.

Sections of focused-refined NLRP3–CRID3 cryo-EM density overlaid with their respective atomic models. Densities are shown as a blue mesh, and sticks are shown for the structure model coloured as in Fig. 2a. The labels ‘\*’ and ‘†’ refer to the ‘multibody refine monomer’ and ‘best decamer’ density, respectively.

### Extended Data Fig. 6 Conformational transitions in NLRP3.

**a**, Overlay of NLRP3 (7PZC) from the decamer structure with the NLRP3–NEK7 complex structure (6NPY)<sup>15</sup>. The proteins were aligned to the NBD-

HD1 subdomains. Only the last three repeats of the LRRs are coloured for clarity. The C-terminal lobe of NEK7 (blue) adopts the space in the concave site of the LRR that is occupied in the NLRP3 decamer structure by the acidic loop (red). **b**, Overlay of NLRP3 from the decamer assembly with NLR family proteins NOD2 (5IRN)<sup>18</sup>, monomeric NLRC4 (4KXF)<sup>19</sup>, and the disc-like NLRC4 (3JBL)<sup>48</sup> structure. **c**, Amino acid register shift in HD2 and trLRR between the NLRP3 decamer and the NLRP3–NEK7 structure. Overlay of residues 435–828 between our structure (7PZC, dark grey/coloured) and the NLRP3–NEK7 structure (6NPY, light grey)<sup>15</sup> showing WHD, HD1, trLRR and 3 repeats of the cnLRR. There is a shift in the amino acid register starting from position E538 that varies between -12 residues in the first helix of HD2 and +44 residues in the first β-strand of trLRR. The register synchronizes again at L737 at the beginning of the cnLRR domain. Segments are coloured according to the amino acid register shifts.

### Extended Data Fig. 7 Arrangements of the trLRR and the cnLRR.

**a**, Sequence alignment of individual repeats of the LRR domain. The trLRR starts at position F650 with the FXXIXI motif and a 26-aa repeat. An LRR-mismatching region of 42 residues interrupts the conventional fold from residue F683 on, forming a flexible, acidic loop. A highly charged stretch of 14 residues (689–702, theoretical pI 4.4) with acidic residues at the tip binds into the concave side of the LRR. The cnLRR starts at position L743 and contains 10 repeats of a proto-typic 28/29 residue alteration<sup>49</sup>. Charged residues in the concave surface of the LRR are coloured blue and red. Leucine residues or homologous hydrophobic residues at LRR-defining positions are indicated bold, and cysteines are boxed yellow. Mismatching residues that preclude a cnLRR fold are boxed cyan. Secondary structure elements of a cnLRR fold are indicated at the top. **b**, Electrostatics of the LRR–acidic loop–LRR' interaction. On the left side is the acidic loop interaction in the LRR without the loop (650–1036, Δ683–727) shown. On the right is the interaction of the C-terminal repeat (998'-1036') of the cognate LRR binding into the concave LRR side shown.

## Extended Data Fig. 8 Mapping of pathogenic disease mutations in human NLRP3.

**a**, Structural mapping of all missense mutations listed in the Infevers database<sup>23</sup>. Missense mutations are highlighted as spheres in light blue; validated, pathogenic mutations are marked in red. Although disease mutations are found in the entire protein, a particular accumulation is seen in the NBD of the NACHT domain. **b**, Overview of the 20 validated CAPS-associated pathogenic mutations from the Infevers database<sup>23</sup> with their predicted effect based on the NLRP3 structure. CAPS, Cryopyrin associated periodic syndrome; CINCA, chronic infantile neurological cutaneous articular; FCAS1, familial cold auto-inflammatory syndrome 1; NOMID, neonatal-onset multisystem inflammatory disorder; MWS, Muckle-Wells syndrome. **c**, Nineteen of twenty pathogenic NLRP3 mutations locate to an interface in the NACHT domain that shears upon activation. The proposed change to an ‘open’ conformation would lead to a disruption of the NBD to WHD/HD2 interface by a ~90° rotation in the linker region between HD1 and WHD. The rotation can be seen by the last helix of HD1 (coloured bordeaux) relative to the first helix of WD1 (coloured olive). The pathogenic mutations may disrupt the integrity of the interface between NBD and WHD/HD2, preventing NLRP3 from entering the resting state of the autoinhibited conformation. The CINCA mutation Y861C is the only residue outside the NACHT domain with a validated pathogenic phenotype. As Y861 interacts with the acidic loop, its mutation could affect the dimer formation mediated by interface A.

## Extended Data Fig. 9 Multiple sequence alignment of NLRP3 proteins.

Secondary structure elements are indicated for human NLRP3 as determined here. Subdomain boundaries are labelled and conserved sequence motifs are written italic. Residues in the decamer assembly interfaces A, B and C are labelled with asterisks. Smaller asterisks correspond to 10–50, larger to > 50 Å<sup>2</sup> buried surface area, respectively. Pathogenic disease mutations FCAS1 (familial cold auto-inflammatory syndrome 1), CINCA (chronic infantile neurological cutaneous articular),

MWS (Muckle-Wells syndrome) and CAPS (Cryopyrin associated periodic syndrome) are marked with circles as indicated. Sequences of human (UniProt accession number Q96P20), macaque (B0FPE9), mouse (Q8R4B8), rat (D4A523), and bovine (A6QLE5) NLRP3 proteins were aligned with MultAlin and the secondary structure annotated with ESPript<sup>50</sup>.

### **Extended Data Fig. 10 The CRID3-binding site in NLRP3.**

**a**, CRID3 binds into a deep crevice that is spanned by subdomains NBD, HD1, WHD, HD2 and trLRR. Only the tertiary alcohol group reaches out of this binding cleft. **b**, Close-up of CRID binding to subdomains NBD, HD1 and WHD (left) and HD1, WHD, HD2 and trLRR (right). To visualize the binding sites, the other subdomains in each case were omitted. **c**, Density map around CRID3 displayed at 4 (black), 7 (red) and 10 (magenta) RMSD threshold. **d**, SPR measurements of NLRP3 mutants A228Q, R351T and R578A showed no binding to CRID3. **e**, Mutational analysis of the CRID3-binding interface in NLRP3 shows that all three mutants could not be activated by nigericin in cell-based assays. This suggests that the integrity of the binding site in between the NBD/HD1 and WHD/HD2 subdomains is critical for the activation mechanism of NLRP3. An ASC speck activation assay was used with HEK293T cells stably expressing an ASC-BFP fusion and transfected with a doxycycline-inducible NLRP3-T2A-mCherry construct containing either wild-type (wt) or CRID3-binding interface mutants. Data are mean  $\pm$  SEM of  $n = 3$  independent experiments (ns  $p > 0.05$ , \*  $p < 0.05$ , \*\*\*  $p < 0.0001$ ) (two-way ANOVA with Tukey's multiple comparisons test).

[Source data](#)

## **Supplementary information**

### **Supplementary Information**

This file contains Supplementary Table 1, gel source data for Extended Data Fig. 1b, gel source data for Extended Data Fig. 1h and gel source data for

Extended Data Fig. 2d, e, f.

## **Reporting Summary**

## **Peer Review File**

## **Supplementary Video 1**

**Central density.** Video animation displaying the density in the central cavity of the NLRP3 decamer.

## **Supplementary Video 2**

**NLRP3 decamer.** Video animation displaying a ribbon diagram of the NLRP3 monomer, its formation of an LRR intertwined dimer and the assembly of the decamer.

## **Source data**

### **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. 10**

## **Rights and permissions**

### **Reprints and Permissions**

## **About this article**

## Cite this article

Hochheiser, I.V., Pilsl, M., Hagelueken, G. *et al.* Structure of the NLRP3 decamer bound to the cytokine release inhibitor CRID3. *Nature* **604**, 184–189 (2022). <https://doi.org/10.1038/s41586-022-04467-w>

- Received: 14 July 2021
- Accepted: 18 January 2022
- Published: 03 February 2022
- Issue Date: 07 April 2022
- DOI: <https://doi.org/10.1038/s41586-022-04467-w>

## Share this article

Anyone you share the following link with will be able to read this content:

[Get shareable link](#)

Sorry, a shareable link is not currently available for this article.

[Copy to clipboard](#)

Provided by the Springer Nature SharedIt content-sharing initiative

---

This article was downloaded by **calibre** from <https://www.nature.com/articles/s41586-022-04467-w>

- Article
- [Published: 30 March 2022](#)

# Differential assembly diversifies GABA<sub>A</sub> receptor structures and signalling

- [Andrija Sente](#)<sup>1</sup>,
- [Rooma Desai](#)<sup>2</sup>,
- [Katerina Naydenova](#)<sup>1</sup>,
- [Tomas Malinauskas](#) [ORCID: orcid.org/0000-0002-4847-5529](#)<sup>3</sup>,
- [Youssef Jounaidi](#)<sup>2</sup>,
- [Jonas Miehling](#)<sup>1</sup>,
- [Xiaojuan Zhou](#)<sup>2</sup>,
- [Simonas Masiulis](#)<sup>1</sup> nAff5,
- [Steven W. Hardwick](#) [ORCID: orcid.org/0000-0001-9246-1864](#)<sup>4</sup>,
- [Dimitri Y. Chirgadze](#) [ORCID: orcid.org/0000-0001-9942-0993](#)<sup>4</sup>,
- [Keith W. Miller](#) [ORCID: orcid.org/0000-0002-4105-6639](#)<sup>2</sup> &
- [A. Radu Aricescu](#) [ORCID: orcid.org/0000-0003-3783-1388](#)<sup>1</sup>

*Nature* volume **604**, pages 190–194 (2022)

- 3816 Accesses
- 69 Altmetric
- [Metrics details](#)

## Subjects

- [Cryoelectron microscopy](#)

- [Ion channels in the nervous system](#)

## Abstract

Type A  $\gamma$ -aminobutyric acid receptors (GABA<sub>A</sub>Rs) are pentameric ligand-gated chloride channels that mediate fast inhibitory signalling in neural circuits<sup>1,2</sup> and can be modulated by essential medicines including general anaesthetics and benzodiazepines<sup>3</sup>. Human GABA<sub>A</sub>R subunits are encoded by 19 paralogous genes that can, in theory, give rise to 495,235 receptor types. However, the principles that govern the formation of pentamers, the permutational landscape of receptors that may emerge from a subunit set and the effect that this has on GABAergic signalling remain largely unknown. Here we use cryogenic electron microscopy to determine the structures of extrasynaptic GABA<sub>A</sub>Rs assembled from  $\alpha$ 4,  $\beta$ 3 and  $\delta$  subunits, and their counterparts incorporating  $\gamma$ 2 instead of  $\delta$  subunits. In each case, we identified two receptor subtypes with distinct stoichiometries and arrangements, all four differing from those previously observed for synaptic,  $\alpha$ 1-containing receptors<sup>4,5,6,7</sup>. This, in turn, affects receptor responses to physiological and synthetic modulators by creating or eliminating ligand-binding sites at subunit interfaces. We provide structural and functional evidence that selected GABA<sub>A</sub>R arrangements can act as coincidence detectors, simultaneously responding to two neurotransmitters: GABA and histamine. Using assembly simulations and single-cell RNA sequencing data<sup>8,9</sup>, we calculated the upper bounds for receptor diversity in recombinant systems and *in vivo*. We propose that differential assembly is a pervasive mechanism for regulating the physiology and pharmacology of GABA<sub>A</sub>Rs.

This is a preview of subscription content

## Access options

Subscribe to Nature+

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

\$29.99

monthly

[Subscribe](#)

Subscribe to Journal

Get full journal access for 1 year

\$199.00

only \$3.90 per issue

[Subscribe](#)

All prices are NET prices.

VAT will be added later in the checkout.

Tax calculation will be finalised during checkout.

Buy article

Get time limited or full article access on ReadCube.

\$32.00

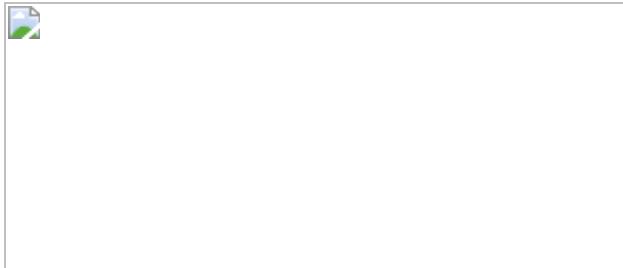
[Buy](#)

All prices are NET prices.

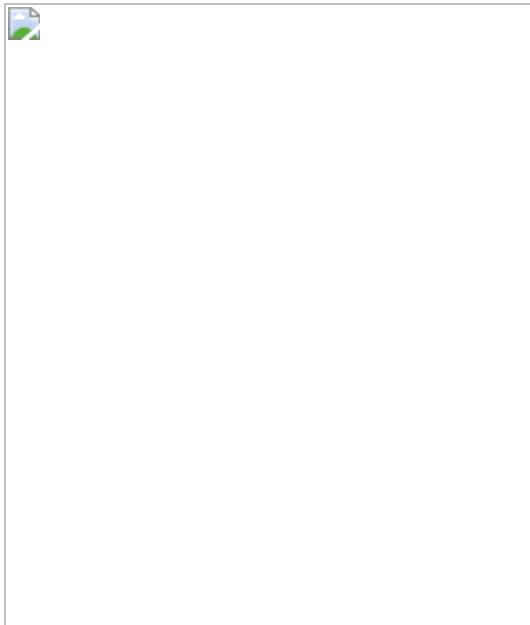
### **Additional access options:**

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

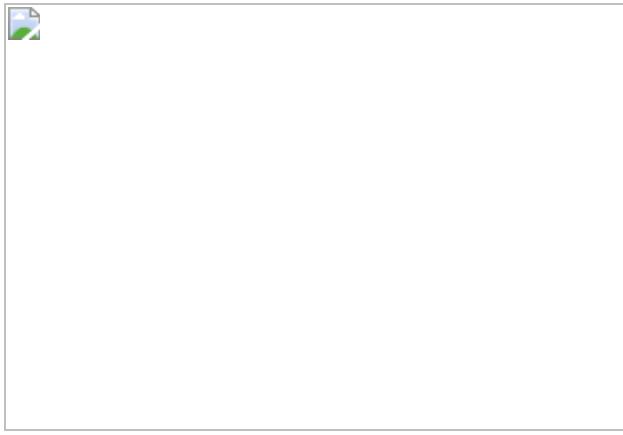
**Fig. 1: Landscape of differential GABA<sub>A</sub> receptor assemblies.**



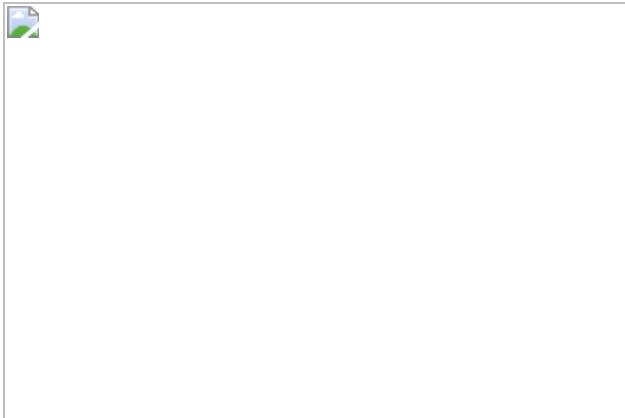
**Fig. 2: Interplay between GABA and histamine at  $\alpha 4\beta 3\delta$  GABA<sub>A</sub>R.**



**Fig. 3: Differential assembly of GABA<sub>A</sub>R affects drug responses.**



**Fig. 4: Computational simulations of receptor assembly.**



## Data availability

Atomic coordinates for the GABA<sub>A</sub>R models reported here have been deposited to the PDB with accession codes [7QN5-9](#) and [7QNA-E](#), and the cryo-EM density maps have been deposited in the Electron Microscopy Data Bank with accession codes [EMD-14067–EMD-14076](#). Raw movies of all datasets have been deposited in the Electron Microscopy Public Image Archive with accession codes [EMPIAR-10909–EMPIAR-10914](#). Coordinate files used during model building can be found in the PDB under the accession codes [7A5V](#) and [6HUG](#). scRNA-seq data used in the study are available from Allen Brain Map: <https://portal.brain-map.org/atlas-and-data/rnaseq/human-multiple-cortical-areas-smart-seq>. [Source data](#) are provided with this paper.

## Code availability

The scripts for scRNA-seq analysis and for running and analysing the receptor assembly simulations are available at [www.github.com/andrijasente/GABA-differential-assembly](https://www.github.com/andrijasente/GABA-differential-assembly).

## References

1. Sigel, E. & Steinmann, M. E. Structure, function, and modulation of GABA<sub>A</sub> receptors. *J. Biol. Chem.* **287**, 40224–40231 (2012).

2. Sieghart, W. & Sperk, G. Subunit composition, distribution and function of GABA<sub>A</sub> receptor subtypes. *Curr. Top. Med. Chem.* **2**, 795–816 (2002).
3. Sieghart, W. & Savic, M. M. International Union of Basic and Clinical Pharmacology. CVI: GABA<sub>A</sub> receptor subtype- and function-selective ligands: key issues in translation to humans. *Pharmacol. Rev.* **70**, 836–878 (2018).
4. Zhu, S. et al. Structure of a human synaptic GABA<sub>A</sub> receptor. *Nature* **559**, 67–72 (2018).
5. Phulera, S. et al. Cryo-EM structure of the benzodiazepine-sensitive  $\alpha 1\beta 1\gamma 2S$  tri-heteromeric GABA<sub>A</sub> receptor in complex with GABA. *eLife* **7**, e39383 (2018).
6. Laverty, D. et al. Cryo-EM structure of the human  $\alpha 1\beta 3\gamma 2$  GABA<sub>A</sub> receptor in a lipid bilayer. *Nature* **565**, 516–520 (2019).
7. Masiulis, S. et al. GABA<sub>A</sub> receptor signalling mechanisms revealed by structural pharmacology. *Nature* **565**, 454–459 (2019).
8. Tasic, B. et al. Shared and distinct transcriptomic cell types across neocortical areas. *Nature* **563**, 72–78 (2018).
9. Hodge, R. D. et al. Conserved cell types with divergent features in human versus mouse cortex. *Nature* **573**, 61–68 (2019).
10. Wisden, W., Laurie, D. J., Monyer, H. & Seuberg, P. H. The distribution of 13 GABA<sub>A</sub> receptor subunit mRNAs in the rat brain. I. Telencephalon, diencephalon, mesencephalon. *J. Neurosci.* **12**, 1040–1062 (1992).
11. Nusser, Z., Sieghart, W. & Somogyi, P. Segregation of different GABA<sub>A</sub> receptors to synaptic and extrasynaptic membranes of cerebellar granule cells. *J. Neurosci.* **18**, 1693–1703 (1998).

12. Pearce, R. A. Physiological evidence for two distinct GABA<sub>A</sub> responses in rat hippocampus. *Neuron* **10**, 189–200 (1993).
13. Baude, A., Sequier, J. M., McKernan, R. M., Olivier, K. R. & Somogyi, P. Differential subcellular distribution of the alpha 6 subunit versus the alpha 1 and beta 2/3 subunits of the GABA<sub>A</sub>/benzodiazepine receptor complex in granule cells of the cerebellar cortex. *Neuroscience* **51**, 739–748 (1992).
14. Farrant, M. & Nusser, Z. Variations on an inhibitory theme: phasic and tonic activation of GABA<sub>A</sub> receptors. *Nat. Rev. Neurosci.* **6**, 215–229 (2005).
15. Brickley, S. G. & Mody, I. Extrasynaptic GABA<sub>A</sub> receptors: their function in the CNS and implications for disease. *Neuron* **73**, 23–34 (2012).
16. Scholze, P. et al. Two distinct populations of  $\alpha 1\alpha 6$ -containing GABA<sub>A</sub>-receptors in rat cerebellum. *Front. Synaptic Neurosci.* **12**, 591129 (2020).
17. Barrera, N. P. et al. Atomic force microscopy reveals the stoichiometry and subunit arrangement of the  $\alpha 4\beta 3\delta$  GABA<sub>A</sub> receptor. *Mol. Pharmacol.* **73**, 960–967 (2008).
18. Patel, B., Mortensen, M. & Smart, T. G. Stoichiometry of delta subunit containing GABA<sub>A</sub> receptors. *Br. J. Pharmacol.* **171**, 985–994 (2014).
19. Wagoner, K. R. & Czajkowski, C. Stoichiometry of expressed  $\alpha_4\beta_2\delta$   $\gamma$ -aminobutyric acid type A receptors depends on the ratio of subunit cDNA transfected. *J. Biol. Chem.* **285**, 14187–14194 (2010).
20. Botzolakis, E. J. et al. Comparison of gamma-aminobutyric acid, type A (GABA<sub>A</sub>), receptor  $\alpha\beta\gamma$  and  $\alpha\beta\delta$  expression using flow cytometry and electrophysiology: evidence for alternative subunit stoichiometries and arrangements. *J. Biol. Chem.* **291**, 20440–20461 (2016).

21. Zhou, X. et al. High-level production and purification in a functional state of an extrasynaptic gamma-aminobutyric acid type A receptor containing  $\alpha 4\beta 3\delta$  subunits. *PLoS ONE* **13**, e0191583 (2018).
22. Eaton, M. M. et al.  $\gamma$ -Aminobutyric acid type A  $\alpha 4$ ,  $\beta 2$ , and  $\delta$  subunits assemble to produce more than one functionally distinct receptor type. *Mol. Pharmacol.* **86**, 647–656 (2014).
23. Nakane, T. et al. Single-particle cryo-EM at atomic resolution. *Nature* **587**, 152–156 (2020).
24. Saras, A. et al. Histamine action on vertebrate GABA<sub>A</sub> receptors: direct channel gating and potentiation of GABA responses. *J. Biol. Chem.* **283**, 10470–10475 (2008).
25. Wongsamitkul, N. et al.  $\alpha$  Subunits in GABA<sub>A</sub> receptors are dispensable for GABA and diazepam action. *Sci. Rep.* **7**, 15498 (2017).
26. Taylor, P. M. et al. Identification of amino acid residues within GABA<sub>A</sub> receptor beta subunits that mediate both homomeric and heteromeric receptor expression. *J. Neurosci.* **19**, 6360–6371 (1999).
27. Bianchi, M. T., Clark, A. G. & Fisher, J. L. The wake-promoting transmitter histamine preferentially enhances alpha-4 subunit-containing GABA<sub>A</sub> receptors. *Neuropharmacology* **61**, 747–752 (2011).
28. Lee, H. J. et al. A pharmacological characterization of GABA, THIP and DS2 at binary  $\alpha 4\beta 3$  and  $\beta 3\delta$  receptors: GABA activates  $\beta 3\delta$  receptors via the  $\beta 3^+\delta^-$  interface. *Brain Res.* **1644**, 222–230 (2016).
29. Miller, P. S. & Aricescu, A. R. Crystal structure of a human GABA<sub>A</sub> receptor. *Nature* **512**, 270–275 (2014).
30. Hoerbelt, P. et al. Mutagenesis and computational docking studies support the existence of a histamine binding site at the extracellular

$\beta 3^+ \beta 3^-$  interface of homooligomeric  $\beta 3$  GABA<sub>A</sub> receptors.  
*Neuropharmacology* **108**, 252–263 (2016).

31. Falch, E. & Krogsgaard-Larsen, P. The binding of the GABA agonist [<sup>3</sup>H]THIP to rat brain synaptic membranes. *J. Neurochem.* **38**, 1123–1129 (1982).
32. Wafford, K. A. & Ebert, B. Gaboxadol—a new awakening in sleep. *Curr. Opin. Pharmacol.* **6**, 30–36 (2006).
33. Hanchar, H. J. et al. Ethanol potently and competitively inhibits binding of the alcohol antagonist Ro15-4513 to  $\alpha 4/\beta 3\delta$  GABA<sub>A</sub> receptors. *Proc. Natl Acad. Sci. USA* **103**, 8546–8551 (2006).
34. Wallner, M., Hanchar, H. J. & Olsen, R. W. Low-dose alcohol actions on  $\alpha 4\beta 3\delta$  GABA<sub>A</sub> receptors are reversed by the behavioral alcohol antagonist Ro15-4513. *Proc. Natl Acad. Sci. USA* **103**, 8540–8545 (2006).
35. Olsen, R. W., Hanchar, H. J., Meera, P. & Wallner, M. GABA<sub>A</sub> receptor subtypes: the “one glass of wine” receptors. *Alcohol* **41**, 201–209 (2007).
36. Linden, A. M. et al. Ro 15-4513 antagonizes alcohol-induced sedation in mice through  $\alpha\beta\gamma 2$ -type GABA<sub>A</sub> receptors. *Front. Neurosci.* **5**, 3 (2011).
37. Storustovu, S. I. & Ebert, B. Pharmacological characterization of agonists at delta-containing GABA<sub>A</sub> receptors: functional selectivity for extrasynaptic receptors is dependent on the absence of  $\gamma 2$ . *J. Pharmacol. Exp. Ther.* **316**, 1351–1359 (2006).
38. Brown, N., Kerby, J., Bonnert, T. P., Whiting, P. J. & Wafford, K. A. Pharmacological characterization of a novel cell line expressing human  $\alpha_4\beta_3\delta$  GABA<sub>A</sub> receptors. *Br. J. Pharmacol.* **136**, 965–974 (2002).

39. Marti-Solano, M. et al. Combinatorial expression of GPCR isoforms affects signalling and drug responses. *Nature* **587**, 650–656 (2020).
40. Rudolph, U. & Knoflach, F. Beyond classical benzodiazepines: novel therapeutic potential of GABA<sub>A</sub> receptor subtypes. *Nat. Rev. Drug Discov.* **10**, 685–697 (2011).
41. Bencsits, E., Ebert, V., Tretter, V. & Sieghart, W. A significant part of native γ-aminobutyric acid<sub>A</sub> receptors containing α4 subunits do not contain γ or δ subunits. *J. Biol. Chem.* **274**, 19613–19616 (1999).
42. Levitan, E. S. et al. Structural and functional basis for GABA<sub>A</sub> receptor heterogeneity. *Nature* **335**, 76–79 (1988).
43. Brickley, S. G., Cull-Candy, S. G. & Farrant, M. Single-channel properties of synaptic and extrasynaptic GABA<sub>A</sub> receptors suggest differential targeting of receptor subtypes. *J. Neurosci.* **19**, 2960–2973 (1999).
44. Mertens, S., Benke, D. & Mohler, H. GABA<sub>A</sub> receptor populations with novel subunit combinations and drug binding profiles identified in brain by α5- and δ-subunit-specific immunopurification. *J. Biol. Chem.* **268**, 5965–5973 (1993).
45. Fritschy, J. M. et al. Five subtypes of type A gamma-aminobutyric acid receptors identified in neurons by double and triple immunofluorescence staining with subunit-specific antibodies. *Proc. Natl Acad. Sci. USA* **89**, 6726–6730 (1992).
46. Duggan, M. J., Pollard, S. & Stephenson, F. A. Immunoaffinity purification of GABA<sub>A</sub> receptor α-subunit iso-oligomers. Demonstration of receptor populations containing α1 α2, α1 α3, and α2 α3 subunit pairs. *J. Biol. Chem.* **266**, 24778–24784 (1991).
47. Owens, D. F. & Kriegstein, A. R. Is there more to GABA than synaptic inhibition? *Nat. Rev. Neurosci.* **3**, 715–727 (2002).

48. Datta, D., Arion, D. & Lewis, D. A. Developmental expression patterns of GABA<sub>A</sub> receptor subunits in layer 3 and 5 pyramidal cells of monkey prefrontal cortex. *Cereb. Cortex* **25**, 2295–2305 (2015).
49. Fatemi, S. H., Folsom, T. D., Rooney, R. J. & Thuras, P. D. Expression of GABA<sub>A</sub> α2-, β1- and ε-receptors are altered significantly in the lateral cerebellum of subjects with schizophrenia, major depression and bipolar disorder. *Transl. Psychiatry* **3**, e303 (2013).
50. Mueller, T. M., Remedies, C. E., Haroutunian, V. & Meador-Woodruff, J. H. Abnormal subcellular localization of GABA<sub>A</sub> receptor subunits in schizophrenia brain. *Transl. Psychiatry* **5**, e612 (2015).
51. Murshudov, G. N. et al. REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr. D Biol. Crystallogr.* **67**, 355–367 (2011).
52. Chen, V. B. et al. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 12–21 (2010).
53. Afonine, P. V. et al. Real-space refinement in PHENIX for cryo-EM and crystallography. *Acta Crystallogr. D Struct. Biol.* **74**, 531–544 (2018).
54. Reeves, P. J., Kim, J. M. & Khorana, H. G. Structure and function in rhodopsin: a tetracycline-inducible system in stable mammalian cell lines for high-level expression of opsin mutants. *Proc. Natl Acad. Sci. USA* **99**, 13413–13418 (2002).
55. Elegheert, J. et al. Lentiviral transduction of mammalian cells for fast, scalable and high-level production of soluble and membrane proteins. *Nat. Protoc.* **13**, 2991–3017 (2018).
56. Dostalova, Z. et al. Human α1β3γ2L γ-aminobutyric acid type A receptors: high-level production and purification in a functional state. *Protein Sci.* **23**, 157–166 (2014).

57. Miller, P. S. et al. Structural basis for GABA<sub>A</sub> receptor potentiation by neurosteroids. *Nat. Struct. Mol. Biol.* **24**, 986–992 (2017).
58. Uchanski, T. et al. Megabodies expand the nanobody toolkit for protein structure determination by single-particle cryo-EM. *Nat. Methods* **18**, 60–68 (2021).
59. Russo, C. J. & Passmore, L. A. Electron microscopy: ultrastable gold substrates for electron cryomicroscopy. *Science* **346**, 1377–1380 (2014).
60. Zheng, S. Q. et al. MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat. Methods* **14**, 331–332 (2017).
61. Rohou, A. & Grigorieff, N. CTFFIND4: fast and accurate defocus estimation from electron micrographs. *J. Struct. Biol.* **192**, 216–221 (2015).
62. Tegunov, D. & Cramer, P. Real-time cryo-electron microscopy data preprocessing with Warp. *Nat. Methods* **16**, 1146–1152 (2019).
63. 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).
64. Asarnow, D., Palovcak, E. & Cheng, Y. asarnow/pyem: UCSF pyem v0.5 (Zenodo, 2019).
65. Scheres, S. H. RELION: implementation of a Bayesian approach to cryo-EM structure determination. *J. Struct. Biol.* **180**, 519–530 (2012).
66. Zivanov, J. et al. New tools for automated high-resolution cryo-EM structure determination in RELION-3. *eLife* **7**, e42166 (2018).
67. Zivanov, J., Nakane, T. & Scheres, S. H. W. Estimation of high-order aberrations and anisotropic magnification from cryo-EM data sets in RELION-3.1. *IUCrJ* **7**, 253–267 (2020).

68. 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).
69. Pettersen, E. F. et al. UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).
70. 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).
71. Kucukelbir, A., Sigworth, F. J. & Tagare, H. D. Quantifying the local resolution of cryo-EM density maps. *Nat. Methods* **11**, 63–65 (2014).
72. Naydenova, K. & Russo, C. J. Measuring the effects of particle orientation to improve the efficiency of electron cryomicroscopy. *Nat. Commun.* **8**, 629 (2017).
73. Goddard, T. D. et al. UCSF ChimeraX: meeting modern challenges in visualization and analysis. *Protein Sci.* **27**, 14–25 (2018).
74. Waterhouse, A. et al. SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Res.* **46**, W296–W303 (2018).
75. Weininger, D. SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules. *J. Chem. Inf. Comput. Sci.* **28**, 31–36 (1988).
76. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* **60**, 2126–2132 (2004).
77. Nicholls, R. A., Fischer, M., McNicholas, S. & Murshudov, G. N. Conformation-independent structural comparison of macromolecules with ProSMART. *Acta Crystallogr. D Biol. Crystallogr.* **70**, 2487–2499 (2014).

78. Forman, S. A. A hydrophobic photolabel inhibits nicotinic acetylcholine receptors via open-channel block following a slow step. *Biochemistry* **38**, 14559–14564 (1999).
79. Sente, A. & Naydenova, K. Scripts for differential assembly diversifies GABA<sub>A</sub> receptor structures and signaling (v1.0.1). Zenodo <https://doi.org/10.5281/zenodo.5884871> (2022).

## Acknowledgements

We thank M. Madan Babu, D. Barford, T. Blundell, R. Calinescu, T. Flock, A. B. Kleist, A. Murzin, D. Malinvern, C. J. Russo, N. Unwin and the members of Aricescu laboratory for helpful discussions and comments on this work; V. B. Kasaragod for help with model building; T. Nakane and K. Yamashita for help with data processing; J. Grimmett and T. Darling for support with scientific computing; S. Chen, G. Cannone, G. Sharov, A. Yeates and B. Ahsan for electron microscopy support; T. Uchanski, E. Pardon and J. Steyaert for the Mb192 megobody. Cryo-EM datasets were collected at the MRC-LMB and Cambridge University Department of Biochemistry EM (BiocEM) facilities. We acknowledge funding from the UK Medical Research Council (MR/L009609/1 and MC\_UP\_1201/15 to A.R.A.), National Institute for General Medical Sciences (1R01-GM135550 to K.W.M. and A.R.A.), the Department of Anesthesia, Critical Care & Pain Medicine at Massachusetts General Hospital (to K.W.M.), MRC-LMB and Cambridge Trust (LMB Cambridge Scholarship to A.S. and K.N.), School of Clinical Medicine, University of Cambridge (MB/PhD fellowship to A.S.), and Boehringer Ingelheim Fonds (PhD Fellowship to J.M.). The cryo-EM facility at the Department of Biochemistry is funded by the Wellcome Trust (206171/Z/17/Z and 202905/Z/16/Z) and the University of Cambridge.

## Author information

### Author notes

1. Simonas Masiulis

Present address: Materials and Structural Analysis Division, Thermo Fisher Scientific, Achtseweg Noord, Eindhoven, Netherlands

## Affiliations

### 1. MRC Laboratory of Molecular Biology, Cambridge, UK

Andrija Sente, Katerina Naydenova, Jonas Miehling, Simona Masiulis & A. Radu Aricescu

### 2. Department of Anesthesia, Critical Care and Pain Medicine, Massachusetts General Hospital, Harvard Medical School, Boston, MA, USA

Rooma Desai, Youssef Jounaidi, Xiaojuan Zhou & Keith W. Miller

### 3. Division of Structural Biology, Wellcome Centre for Human Genetics, University of Oxford, Oxford, UK

Tomas Malinauskas

### 4. Department of Biochemistry, University of Cambridge, Cambridge, UK

Steven W. Hardwick & Dimitri Y. Chirgadze

## Contributions

A.S., K.W.M. and A.R.A. conceived the project. A.S., Y.J. and R.D. generated and characterized cell lines. A.S. and S.M. purified proteins and prepared the cryo-EM samples. A.S., S.W.H. and D.Y.C. collected the cryo-EM data. A.S., K.N. and S.M. processed the cryo-EM data. A.S., T.M., J.M. and A.R.A. built and refined the atomic models. R.D. and K.W.M. designed, performed and analysed the electrophysiology experiments. X.Z. and K.W.M. designed, performed and analysed the radioligand binding assays. A.S. designed and performed the bioinformatics analyses. A.S. and K.N. designed, performed and analysed the simulations for receptor

assembly. A.S., K.N., R.D., K.W.M. and A.R.A. wrote the manuscript, with input from all authors.

## Corresponding authors

Correspondence to [Andrija Sente](#), [Keith W. Miller](#) or [A. Radu Aricescu](#).

## Ethics declarations

### Competing interests

The authors declare no competing interests.

## Peer review

### Peer review information

*Nature* thanks Margot Ernst and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

## Additional information

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## Extended data figures and tables

### [Extended Data Fig. 1 Quality of cryo-EM maps and models.](#)

Each panel contains masked, unmasked, phase-randomised and map vs model Fourier shell correlation (FSC) curves (top), particle orientation distribution (middle) and local resolution-coloured maps for each structure. **a**,  $\alpha 4\beta 3\delta$  "apo" (+HEPES). **b**,  $\beta 3\delta$  "apo" (+HEPES). **c**,  $\alpha 4\beta 3\delta + GABA + histamine$ . **d**,  $\beta 3\delta + histamine$ . **e**,  $\alpha 4\beta 3\delta + histamine$ . **f**,  $\alpha 4\beta 3\gamma 2 + GABA$ . **g**,

$\beta 3\gamma 2 + \text{GABA}$ . **h**,  $\alpha 4\beta 3\delta + \text{THIP} + \text{histamine}$ . **i**,  $\beta 3\delta + \text{THIP} + \text{histamine}$ . **j**,  $\alpha 1\beta 3\gamma 2 + \text{Ro15-4513}$ . The probability density function for a certain orientation ranges between 0 (blue, low) and 0.0001 (red, high).

### Extended Data Fig. 2 Survey of ligands bound in agonist pockets in different datasets.

**a**,  $\alpha 4\beta 3\delta$  and  $\beta 3\delta$  GABA<sub>A</sub>R solved in the absence of any known modulators. In both subtypes, HEPES is bound in the  $\beta 3^+/\beta 3^-$  agonist pocket under loop C. **b**,  $\alpha 4\beta 3\delta$  and  $\beta 3\delta$  GABA<sub>A</sub>R solved in the presence of 0.2 mM GABA and 1 mM histamine. **c**,  $\alpha 4\beta 3\gamma 2$  and  $\beta 3\gamma 2$  GABA<sub>A</sub>R solved in the presence of 0.2 mM GABA. **d**,  $\alpha 4\beta 3\delta$  and  $\beta 3\delta$  GABA<sub>A</sub>R solved in the presence of 1 mM THIP and 1 mM histamine. **e**, Summary table illustrating ligand vs interface compatibility. Empty spaces are left in cases where insufficient data is available to make a conclusion. For the GABA binding to  $\beta 3^+/\gamma 2^-$  pocket, a question mark indicates variable occupancy. **f**, Comparison of pore profiles of  $\alpha 4\beta 3\delta + \text{GABA} + \text{histamine}$ ,  $\alpha 4\beta 3\delta + \text{histamine}$ ,  $\alpha 1\beta 3\gamma 2 + \text{alprazolam}$  (6HUO) and  $\alpha 1\beta 3\gamma 2 + \text{bicuculline}$  (6HUK).

### Extended Data Fig. 3 Ligand coordination in the “agonist pockets” under the loops C at various inter-subunit interfaces.

All pockets in which a ligand was observed in our structures are shown with a single representative example. Potential hydrogen bonds are displayed as dashed yellow lines. Histamine and GABA in the  $\beta 3^+/\alpha 4^-$  pocket are taken from the  $\alpha 4\beta 3\delta + \text{GABA} + \text{histamine}$  structure. THIP in the  $\beta 3^+/\alpha 4^-$  and  $\delta^+/\beta 3^-$  pocket is taken from the  $\alpha 4\beta 3\delta + \text{THIP} + \text{histamine}$  structure. Histamine in the  $\beta 3^+/\beta 3^-$  pocket is taken from the  $\alpha 4\beta 3\delta + \text{histamine}$  structure. HEPES in the  $\beta 3^+/\beta 3^-$  pocket is taken from the "apo"  $\alpha 4\beta 3\delta$  structure. GABA in the  $\beta 3^+/\gamma 2^-$  pocket is taken from the  $\alpha 4\beta 3\gamma 2 + \text{GABA}$  structure. Chain IDs represented in the panels are given as capital letters next to the subunit type.

## Extended Data Fig. 4 Characterizing GABA and histamine effects on GABA<sub>A</sub> receptors composed of α4, β3 and δ subunits.

**a–c**, Unlike α4β3δ receptors (**a**), β3δ (**b**) and β3 (**c**) receptors are not gated by 200 μM GABA. Representative current traces from three successive pulses separated by 1 min: (1) 4 s pulse of 10 mM GABA plus 30 μM Etomidate (total current, purple); (2) 8 s pulse of 200 μM GABA (green), and (3) repeat of pulse (1) (pink). These pulses are applied to: **a**, cells co-expressing α4, β3 and δ subunits; **b**, cells co-expressing β3 and δ subunits; **c**, cells expressing the β3 subunit alone. **d, e**, Representative currents normalized to 10 mM GABA currents in the same cell were obtained with varying concentration of (**a**) GABA (0.001–10 μM, n = 38 cells with 3–8 cells for each concentration) or (**b**) histamine (0.03–10 mM, n = 31 cells with 3–8 cells for each concentration). In panel **e**, orange dashed line denotes the baseline. **f**, Hill plot of peak amplitude obtained with the normalized currents from (**a**) and (**b**) plotted against agonist concentration. The EC<sub>50</sub> for GABA was 69.3 nM (95% CI 40.5–118.3 nM) and for histamine was 821.2 μM (95% CI 340.5–1981.0 μM) and Hill coefficient of 0.9 (95% CI 0.5–1.3) and 1.1 (95% CI 0.3–2.0) respectively. **g**, Scatter plot of peak amplitudes of currents shown in Fig. [2d](#) by GABA (10 mM, n = 7 cells) and histamine (10 mM, n = 6 cells) alone or co-applied (n = 8 cells) in the first pulse as a percent of currents elicited by 10 mM GABA + 30 μM Etomidate. One-way ANOVA P < 0.0001. **h–k**, Scatter plots of the time constants ( $\tau$ ) and amplitudes for the slow and fast phases of deactivation that were obtained by fitting a double exponential to the falling phase of currents such as those in Fig. [2e](#) (n = 4 cells for each experiments). In every paired experiment, the direction of change was consistent. In the presence of histamine fast and slow time constants decreased significantly (P = 0.03, two-tailed paired t test) and the fraction of receptors with slow deactivation increased (P = 0.04, two-tailed paired t test). Significance levels compared to controls were determined by ordinary one-way ANOVA or by the Paired Student's t test and are denoted as \*\*\* P < 0.0001 and \* P < 0.05. Data are presented as mean ± SD.

[Source data](#)

## Extended Data Fig. 5 GABA and histamine crosstalk happens at the level of individual $\alpha$ 4 $\beta$ 3 $\delta$ receptors.

To address the complexity arising from the presence of multiple receptor subtypes, we investigated the modulation of 100 nM GABA currents by 300  $\mu$ M histamine in cells expressing different combinations of  $\alpha$ 4,  $\beta$ 3 and  $\delta$  subunits. We first established that 100 nM GABA robustly activates currents in  $\alpha$ 4 $\beta$ 3 $\delta$  cells ( $n = 6$  cells) (a), barely in  $\alpha$ 4 $\beta$ 3 ( $n = 5$  cells) (b) and not in  $\beta$ 3 $\delta$  ( $n = 6$  cells) (c) or  $\beta$ 3 cells ( $n = 4$  cells) (d). Histamine at 300  $\mu$ M robustly enhances the GABA current in  $\alpha$ 4 $\beta$ 3 $\delta$  cells but only modestly in the  $\alpha$ 4 $\beta$ 3,  $\beta$ 3 $\delta$  and  $\beta$ 3 cells. The modest enhancement is attributable to histamine currents alone (e.g., from the  $\beta$ 3 $\delta$  or  $\beta$ 3-homomeric subtypes). Representative current traces from three successive pulses separated by 1 min: (1) 8 s pulse of 100 nM GABA (dark green); (2) 8 s pulse of 100 nM GABA with a 4 s notch of 300  $\mu$ M histamine added after two seconds (blue), and (3) 8 s pulse of 100 nM GABA (light green). Finally, to open all receptor isoforms, a normalization pulse of (10 mM GABA + 30  $\mu$ M etomidate) was applied. e, Histamine enhances GABA current strongly only in  $\alpha$ 4 $\beta$ 3 $\delta$  receptors ( $P < 0.0001$ , one-way ANOVA). Scatter plot shows peak current amplitudes of the blue traces in panels a–d, normalized to the total cell current (10 mM GABA + 30  $\mu$ M etomidate). Data are presented as mean  $\pm$  SD. Sample size is the same as in panels a–d.

[Source data](#)

## Extended Data Fig. 6 Ro15-4513 cannot bind to the $\alpha$ 4 $^+$ / $\delta$ $^-$ interface.

a, b, Two views of Ro15-4513 bound in the  $\alpha$ 1 $^+/\gamma$ 2 $^-$  benzodiazepine pocket of an  $\alpha$ 1 $\beta$ 3 $\gamma$ 2 receptor. c, d, Equivalent views of an  $\alpha$ 4 $\beta$ 3 $\delta$  receptor show that clashes with the side chains of R135 ( $\alpha$ 4 $^+$ ), E71 and H92 ( $\delta$  $^-$ ) prevent Ro15-4513 (and, likely, any benzodiazepine ligand) from binding to the  $\alpha$ 4 $^+/\delta$  $^-$  interface in the same mode as observed in  $\alpha$ 1 $\beta$ 3 $\gamma$ 2.

## Extended Data Fig. 7 Single-cell expression patterns of GABA<sub>A</sub>R subunits in the human cortex.

**a**, Grouping cortical cell types by their GABA<sub>A</sub>R subunit expression. The majority of cell types (31) expresses a combination of 11 different GABA<sub>A</sub>R subunits, including  $\alpha$ 1-5,  $\beta$ 1-3,  $\delta$  and  $\gamma$ 2-3. Cell type "Exc L5-6 THEMIS THTPA" expresses the highest number of subunits (14):  $\alpha$ 1-5,  $\beta$ 1-3,  $\delta$ ,  $\varepsilon$ ,  $\gamma$ 1-3 and  $\rho$ 1. Expression in this panel is binarized such that any subunit with a trimmed mean of summed intron and exon counts larger than 0 count per million (CPM) is considered as expressed ([Methods](#)). **b**, Heat map showing trimmed means of CPMs for every GABA<sub>A</sub>R subunit across individual cell types in the human cortex. Cell types are clustered by their gene expression profiles.

## Extended Data Fig. 8 Cryo-EM data processing strategy.

**a**, A typical cryo-EM data processing workflow. The numbers in this case are derived from the  $\alpha$ 4 $\beta$ 3 $\delta$  "apo" (+HEPES) dataset. **b**, Soft mask used to separate receptors with 2 or 3 Nb25 bound, which largely (but not completely) separates  $\alpha$ 4 $\beta$ 3 $\delta$  and  $\beta$ 3 $\delta$  receptors. The separation is incomplete in all cases because Nb25 may also bind weakly to the  $\beta$ 3<sup>+</sup>/ $\alpha$ 4<sup>-</sup> interface. **c**, To further separate  $\alpha$ 4 $\beta$ 3 $\delta$  and  $\beta$ 3 $\delta$  receptors, a soft mask around the putative position of the vestibule glycan on the  $\alpha$ 4 subunit and the putative position of the N149 glycan on the  $\beta$ 3 subunit is used. **d**, Soft mask around the  $\alpha$ 4 vestibule glycan used to separate  $\alpha$ 4 $\beta$ 3 $\gamma$ 2 from  $\beta$ 3 $\gamma$ 2 receptor subtypes.

**Extended Data Table 1 Cryo-EM data collection parameters**

**Extended Data Table 2 Refinement and validation statistics**

## **Supplementary information**

### Supplementary Information

This file contains Supplementary Figures; Supplementary Tables; Supplementary Methods; Supplementary Discussion and Supplementary References

## **Reporting Summary**

## **Peer Review File**

## **Source data**

### **Source Data Fig. 2**

### **Source Data Fig. 3**

### **Source Data Extended Data Fig. 4**

### **Source Data Extended Data Fig. 5**

### **Source Data Extended Data Fig. 7**

## **Rights and permissions**

### **Reprints and Permissions**

## **About this article**

### **Cite this article**

Sente, A., Desai, R., Naydenova, K. *et al.* Differential assembly diversifies GABA<sub>A</sub> receptor structures and signalling. *Nature* **604**, 190–194 (2022).  
<https://doi.org/10.1038/s41586-022-04517-3>

- Received: 10 August 2021

- Accepted: 04 February 2022
- Published: 30 March 2022
- Issue Date: 07 April 2022
- DOI: <https://doi.org/10.1038/s41586-022-04517-3>

## Share this article

Anyone you share the following link with will be able to read this content:

[Get shareable link](#)

Sorry, a shareable link is not currently available for this article.

[Copy to clipboard](#)

Provided by the Springer Nature SharedIt content-sharing initiative

---

This article was downloaded by **calibre** from <https://www.nature.com/articles/s41586-022-04517-3>

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

- Article
- [Published: 30 March 2022](#)

# Structure, substrate recognition and initiation of hyaluronan synthase

- [Finn P. Maloney<sup>1</sup>](#)<sup>na1</sup>,
- [Jeremi Kuklewicz](#) [ORCID: orcid.org/0000-0002-2204-8844](#)<sup>1</sup><sup>na1</sup>,
- [Robin A. Corey](#)<sup>2</sup>,
- [Yunchen Bi](#)<sup>3,8</sup>,
- [Ruoya Ho](#) [ORCID: orcid.org/0000-0002-2369-8443](#)<sup>1</sup>,
- [Lukasz Mateusiak](#) [ORCID: orcid.org/0000-0002-2837-8342](#)<sup>4</sup>,
- [Els Pardon](#) [ORCID: orcid.org/0000-0002-2466-0172](#)<sup>5,6</sup>,
- [Jan Steyaert](#) [ORCID: orcid.org/0000-0002-3825-874X](#)<sup>5,6</sup>,
- [Phillip J. Stansfeld](#) [ORCID: orcid.org/0000-0001-8800-7669](#)<sup>7</sup> &
- [Jochen Zimmer](#) [ORCID: orcid.org/0000-0002-8423-2882](#)<sup>1</sup>

*Nature* volume **604**, pages 195–201 (2022)

- 4442 Accesses
- 71 Altmetric
- [Metrics details](#)

## Subjects

- [Cryoelectron microscopy](#)
- [Glycobiology](#)
- [Polysaccharides](#)

# Abstract

Hyaluronan is an acidic heteropolysaccharide comprising alternating *N*-acetylglucosamine and glucuronic acid sugars that is ubiquitously expressed in the vertebrate extracellular matrix<sup>1</sup>. The high-molecular-mass polymer modulates essential physiological processes in health and disease, including cell differentiation, tissue homeostasis and angiogenesis<sup>2</sup>. Hyaluronan is synthesized by a membrane-embedded processive glycosyltransferase, hyaluronan synthase (HAS), which catalyses the synthesis and membrane translocation of hyaluronan from uridine diphosphate-activated precursors<sup>3,4</sup>. Here we describe five cryo-electron microscopy structures of a viral HAS homologue at different states during substrate binding and initiation of polymer synthesis. Combined with biochemical analyses and molecular dynamics simulations, our data reveal how HAS selects its substrates, hydrolyses the first substrate to prime the synthesis reaction, opens a hyaluronan-conducting transmembrane channel, ensures alternating substrate polymerization and coordinates hyaluronan inside its transmembrane pore. Our research suggests a detailed model for the formation of an acidic extracellular heteropolysaccharide and provides insights into the biosynthesis of one of the most abundant and essential glycosaminoglycans in the human body.

This is a preview of subscription content

## Access options

Subscribe to Nature+

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

\$29.99

monthly

[Subscribe](#)

Subscribe to Journal

Get full journal access for 1 year

\$199.00

only \$3.90 per issue

[Subscribe](#)

All prices are NET prices.

VAT will be added later in the checkout.

Tax calculation will be finalised during checkout.

Buy article

Get time limited or full article access on ReadCube.

\$32.00

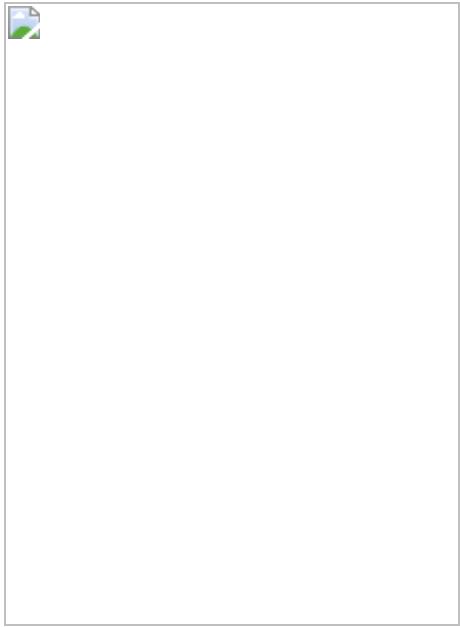
[Buy](#)

All prices are NET prices.

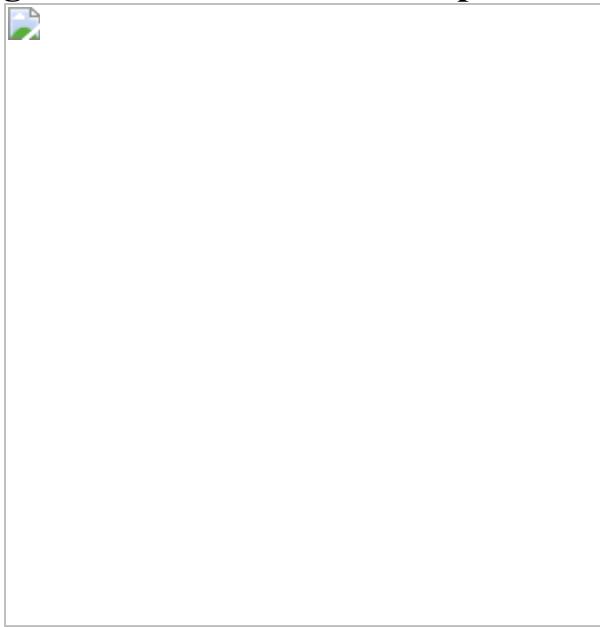
### **Additional access options:**

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

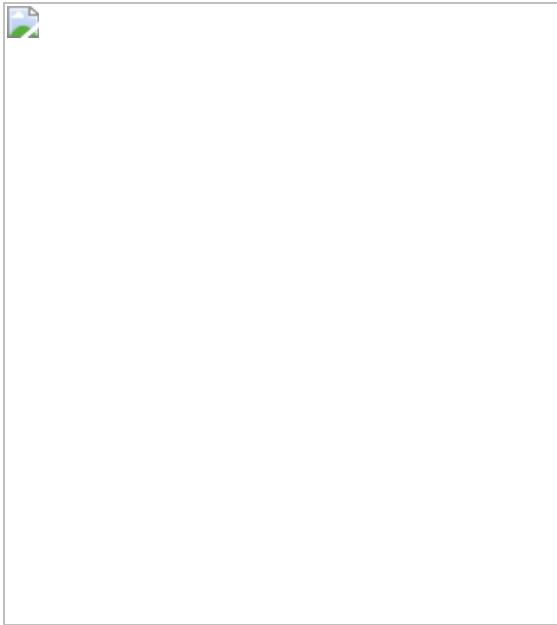
**Fig. 1: The structure of HAS.**



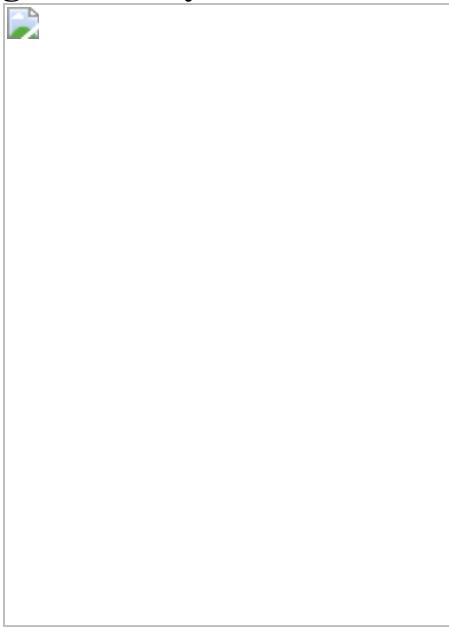
**Fig. 2: Substrate-bound and primed Cv-HAS conformations.**



**Fig. 3: Structural rearrangements after GlcNAc priming.**



**Fig. 4: The dynamics of HA-bound HAS.**



## Data availability

Raw EM videos and maps have been deposited at the Protein Data Bank and Electron Microscopy Data Bank under accession codes [7SP7](#) and [EMD-25367](#); [7SP6](#) and [EMD-25366](#); [7SP8](#) and [EMD-25368](#); [7SP9](#) and [EMD-25369](#); and [7SPA/EMD-25370](#) for the UDP-bound, D302N apo,

UDP-GlcNAc-bound, primed (closed) and primed (open) states, respectively.

## References

1. Girish, K. S. & Kemparaju, K. The magic glue hyaluronan and its eraser hyaluronidase: a biological overview. *Life Sci.* **80**, 1921–1943 (2007).
2. Cyphert, J. M., Trempus, C. S. & Garantziotis, S. Size matters: molecular weight specificity of hyaluronan effects in cell biology. *Int. J. Cell Biol.* **2015**, 563818 (2015).
3. Vigetti, D. et al. Hyaluronan: biosynthesis and signaling. *Biochim. Biophys. Acta* **1840**, 2452–2459 (2014).
4. DeAngelis, P. Hyaluronan synthases: fascinating glycosyltransferases from vertebrates, bacterial pathogens, and algal viruses. *Cell. Mol. Life Sci.* **56**, 670–682 (1999).
5. Sironen, R. et al. Hyaluronan in human malignancies. *Exp. Cell. Res.* **317**, 383–391 (2011).
6. Cowman, M. K., Lee, H.-G., Schwertfeger, K. L., McCarthy, J. B. & Turley, E. A. The content and size of hyaluronan in biological fluids and tissues. *Front. Immunol.* **6**, 261 (2015).
7. Hubbard, C., McNamara, J., Azumaya, C., Patel, M. & Zimmer, J. The hyaluronan synthase catalyzes the synthesis and membrane translocation of hyaluronan. *J. Mol. Biol.* **418**, 21–31 (2012).
8. Weigel, P. H. & Deangelis, P. L. Hyaluronan synthases: a decade-plus of novel glycosyltransferases. *J. Biol. Chem.* **282**, 36777–36781 (2007).
9. DeAngelis, P. L., Jing, W., Drake, R. R. & Achyuthan, A. M. Identification and molecular cloning of a unique hyaluronan synthase from *Pasteurella multocida*. *J. Biol. Chem.* **273**, 8454–8458 (1998).

10. Camenisch, T. D. et al. Disruption of hyaluronan synthase-2 abrogates normal cardiac morphogenesis and hyaluronan-mediated transformation of epithelium to mesenchyme. *J. Clin. Invest.* **106**, 349–360 (2000).
11. Itano, N. et al. Three isoforms of mammalian hyaluronan synthases have distinct enzymatic properties. *J. Biol. Chem.* **274**, 25085–25092 (1999).
12. DeAngelis, P., Jing, W., Graves, M., Burbank, D. & Van Etten, J. Hyaluronan synthase of chlorella virus PBCV-1. *Science* **278**, 1800–1803 (1997).
13. Blackburn, M. R. et al. Distinct reaction mechanisms for hyaluronan biosynthesis in different kingdoms of life. *Glycobiology* **28**, 108–121 (2018).
14. Pardon, E. et al. A general protocol for the generation of nanobodies for structural biology. *Nat. Protoc.* **9**, 674–693 (2014).
15. Tlapak-Simmons, V. L., Baron, C. A. & Weigel, P. H. Characterization of the purified hyaluronan synthase from *Streptococcus equisimilis*. *Biochemistry* **43**, 9234–9242 (2004).
16. 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).
17. Lairson, L. L., Henrissat, B., Davies, G. J. & Withers, S. G. Glycosyltransferases: structures, functions, and mechanisms. *Annu. Rev. Biochem.* **77**, 521–555 (2008).
18. Cantarel, B., Coutinho, P., Rancurel, C. & Bernard, T. The Carbohydrate-Active EnZymes database (CAZy): an expert resource for glycogenomics. *Nucleic Acids Res.* **37**, D233–D238 (2009).
19. Wu, Q. et al. Protein contact prediction using metagenome sequence data and residual neural networks. *Bioinformatics* **36**, 41–48 (2020).

20. Tunyasuvunakool, K., Adler, J., Wu, Z., Jumper, J. & Hassabis, D. Highly accurate protein structure prediction for the human proteome. *Nature* **596**, 590–596 (2021).
21. Baek, M. et al. Accurate prediction of protein structures and interactions using a three-track neural network. *Science* **373**, 871–876 (2021).
22. Charnock, S. J. & Davies, G. J. Structure of the nucleotide-diphospho-sugar transferase, SpsA from *Bacillus subtilis*, in native and nucleotide-complexed forms. *Biochemistry* **38**, 6380–6385 (1999).
23. Morgan, J., Strumillo, J. & Zimmer, J. Crystallographic snapshot of cellulose synthesis and membrane translocation. *Nature* **493**, 181–186 (2013).
24. Purushotham, P., Ho, R. & Zimmer, J. Architecture of a catalytically active homotrimeric plant cellulose synthase complex. *Science* **369**, 1089–1094 (2020).
25. Morgan, J. L. W., McNamara, J. T. & Zimmer, J. Mechanism of activation of bacterial cellulose synthase by cyclic di-GMP. *Nat. Struct. Mol. Biol.* **21**, 489–496 (2014).
26. McManus, J., Yang, H., Wilson, L., Kubicki, J. & Tien, M. Initiation, elongation, and termination of bacterial cellulose synthesis. *ACS Omega* **3**, 2690–2698 (2018).
27. Orlean, P. & Funai, D. Priming and elongation of chitin chains: Implications for chitin synthase mechanism. *Cell Surf.* **5**, 100017 (2019).
28. Yang, J. et al. Key role of the carboxyl terminus of hyaluronan synthase in processive synthesis and size control of hyaluronic acid polymers. *Biomacromolecules* **18**, 1064–1073 (2017).
29. Morgan, J. L. et al. Observing cellulose biosynthesis and membrane translocation in crystallo. *Nature* **531**, 329–334 (2016).

30. PyMol: the PyMOL molecular graphics system v.2.5.0 (Schrödinger, 2021).
31. Ho, B. & Gruswitz, F. HOLLOW: generating accurate representations of channel and interior surfaces in molecular structures. *BMC Struct. Biol.* **8**, 49 (2008).
32. Studier, F. Protein production by auto-induction in high density shaking cultures. *Protein Expr. Purif.* **41**, 207–234 (2005).
33. Kelly, S. J., Taylor, K. B., Li, S. & Jedrzejas, M. J. Kinetic properties of *Streptococcus pneumoniae* hyaluronate lyase. *Glycobiology* **11**, 297–304 (2001).
34. Rohou, A. & Grigorieff, N. CTFFIND4: fast and accurate defocus estimation from electron micrographs. *J. Struct. Biol.* **192**, 216–221 (2015).
35. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D* **60**, 2126–2132 (2004).
36. Adams, P. et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D* **66**, 213–221 (2010).
37. Pettersen, E. F. et al. UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).
38. Huang, J. et al. CHARMM36m: an improved force field for folded and intrinsically disordered proteins. *Nature* **14**, 71–73 (2016).
39. Jo, S., Kim, T., Iyer, V. G. & Im, W. CHARMM-GUI: a web-based graphical user interface for CHARMM. *J. Comput. Chem.* **29**, 1859–1865 (2008).
40. Lee, J. et al. CHARMM-GUI input generator for NAMD, GROMACS, AMBER, OpenMM, and CHARMM/OpenMM simulations using the

CHARMM36 additive force field. *J. Chem. Theo. Comput.* **12**, 405–413 (2016).

41. Sondergaard, C. R., Olsson, M. H. M., Rostkowski, M. & Jensen, J. H. Improved treatment of ligands and coupling effects in empirical calculation and rationalization of  $pK_a$  values. *J. Chem. Theory Comput.* **7**, 2284–2295 (2011).
42. Olsson, M. H. M., Sondergaard, C. R., Rostkowski, M. & Jensen, J. H. PROPKA3: consistent treatment of internal and surface residues in empirical  $pK_a$  predictions. *J. Chem. Theory Comput.* **7**, 525–537 (2011).
43. Bussi, G., Donadio, D. & Parrinello, M. Canonical sampling through velocity rescaling. *J. Chem. Phys.* **126**, 014101 (2007).
44. Parrinello, M. & Rahman, A. Polymorphic transitions in single crystals: a new molecular dynamics method. *J. Appl. Phys.* **52**, 7182–7190 (1981).
45. Berendsen, H. J. C., van der Spoel, D. & van Drunen, R. GROMACS: a message-passing parallel molecular dynamics implementation. *Comput. Phys. Commun.* **91**, 43–56 (1994).
46. Humphrey, W., Dalke, A. & Schulten, K. VMD: visual molecular dynamics. *J. Mol. Graph.* **14**, 33–38 (1996).
47. 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).
48. Harris, C. R. et al. Array programming with NumPy. *Nature* **585**, 357–362 (2020).
49. Hunter, J. D. Matplotlib: a 2D graphics environment. *Comput. Sci. Eng.* **9**, 90–95 (2007).

50. Bernsel, A., Viklund, H., Hennerdal, A. & Elofsson, A. TOPCONS: consensus prediction of membrane protein topology. *Nucleic Acids Res.* **37**, W465–W468 (2009).

## Acknowledgements

We thank K. Dryden and M. Purdy from the MEMC at the University of Virginia as well as A. Wier and the support staff at the NCI-NCEF. E.P. and J.S. acknowledge the support of Instruct-ERIC, part of the European Strategy Forum on Research Infrastructures (ESFRI), and the Research Foundation–Flanders (FWO) for supporting the nanobody discovery and thank E. Beke for technical assistance. J.Z. and F.P.M. were supported by NIH grant R21AI148853. R.H. was supported by NIH grant R01GM101001 (awarded to J.Z.). R.A.C. and P.J.S. are supported by Wellcome (208361/Z/17/Z). P.J.S.’s laboratory is supported by awards from the BBSRC (BB/P01948X/1, BB/R002517/1 and BB/S003339/1) and the MRC (MR/S009213/1). P.J.S. acknowledges the University of Warwick Scientific Computing Research Technology Platform for computational access.

## Author information

### Author notes

1. These authors contributed equally: Finn P. Maloney, Jeremi Kuklewicz

## Affiliations

1. Department of Molecular Physiology and Biological Physics,  
University of Virginia School of Medicine, Charlottesville, VA, USA

Finn P. Maloney, Jeremi Kuklewicz, Ruoya Ho & Jochen Zimmer

2. Department of Biochemistry, University of Oxford, Oxford, UK

Robin A. Corey

3. Laboratory for Marine Biology and Biotechnology, Pilot National Laboratory for Marine Science and Technology (Qingdao), Qingdao, China

Yunchen Bi

4. Laboratory for In Vivo Cellular and Molecular Imaging, ICMI-BEFY, Vrije Universiteit Brussel, Brussels, Belgium

Lukasz Mateusiak

5. VIB-VUB Center for Structural Biology, VIB, Brussels, Belgium

Els Pardon & Jan Steyaert

6. Structural Biology Brussels, Vrije Universiteit Brussel, VUB, Brussels, Belgium

Els Pardon & Jan Steyaert

7. School of Life Sciences and Department of Chemistry, University of Warwick, Coventry, UK

Phillip J. Stansfeld

8. CAS and Shandong Province Key Laboratory of Experimental Marine Biology, Institute of Oceanology, Center for Ocean Mega-Science, Chinese Academy of Sciences, Qingdao, China

Yunchen Bi

## Contributions

Y.B. cloned Cv-HAS and generated materials for nanobody production. E.P. and J.S. generated the nanobodies. Y.B. and L.M. characterized the nanobodies. L.M. performed thermostability assays. F.P.M. purified Cv-HAS–nanobody complexes and produced nanodiscs. R.H. collected EM data, and F.P.M. and J.K. processed the data. F.P.M. determined the apo and

UDP-bound Cv-HAS structures. J.K. determined the substrate-bound and primed Cv-HAS structures. R.A.C. performed all MD simulations. J.Z. wrote the first manuscript. F.P.M., J.K., R.A.C., P.J.S. and J.Z. edited the draft. All of the authors commented on the manuscript.

## Corresponding author

Correspondence to [Jochen Zimmer](#).

## Ethics declarations

## Competing interests

The authors declare no competing interests.

## Peer review

## Peer review information

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

## Additional information

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## Extended data figures and tables

### [Extended Data Fig. 1 Sequence alignment of HAS orthologues.](#)

Comparison of HAS primary sequences from *Chlorella virus* (Cv), *Homo sapiens* (Hs) and *Streptococcus equisimilis* (Se). Topology predictions were

performed using TopCons<sup>50</sup>. Cylinders indicate secondary structure elements observed in Cv-HAS.

### Extended Data Fig. 2 Identification, data collection, and processing of Cv-HAS bound to two nanobodies and UDP.

**(a)** Increased melting temperature of Cv-HAS in the presence of Nb872. Protein melting was measured based on enzymatic activity detected by quantifying the release of UDP in real time. **(b)** HA biosynthesis in the presence of the indicated nanobodies and based on quantification of <sup>3</sup>H-labelled HA by scintillation counting. Data is normalized relative to product yields in the absence of nanobodies. Error bars represent deviations from the means with n = 3 independent experiments. **(c)** Representative autoradiography of <sup>14</sup>C-labelled HA produced in the presence of the indicated nanobodies. The experiment has been repeated at least 4 times with essentially identical results. NC: Negative control in the absence of UDP-GlcNAc substrate (for panel b) or UDP-GlcA (for panel c). PC: Positive control in the absence of nanobody. Lyase: Hyaluronan lyase treatment prior to SDS-PAGE. **(d)** This workflow produced the UDP-bound Cv-HAS structure.

### Extended Data Fig. 3 Cryo-EM data collection and processing of Cv-HAS D302N in the presence of substrate.

This workflow generated the apo, substrate-bound, primed, and primed with open channel Cv-HAS structures.

### Extended Data Fig. 4 Map quality and model building of UDP-bound Cv-HAS.

**(a-d)** Map overview, estimated resolution based on FSC, and particle orientation distribution. **(e)** Secondary structure elements and topology of Cv-HAS. **(f-j)** TM helices 2 to 6 of Cv-HAS. **(k)** TMH3-4 extracellular loop. **(l)** The extracellular TMH5-6 loop. **(m)** The QxxRW motif. **(n)** The C-terminal cytosolic helix. **(o)** The unresolved TMH5-IF3 loop. All maps are contoured at 7.0σ.

## Extended Data Fig. 5 Predicted location of TMH1.

**(a)** Relationship of evolutionarily coupled residues within Cv-HAS' TM and GT regions, generated in MapPred based on 65,535 sequences. TMH1 is shown at its predicted location as a violet cylinder. **(b)** RoseTTAfold models of full-length Cv-HAS. Cv-HAS is shown as a surface and its TMH 2 as a blue cylinder. TMH1 is shown as a cartoon at its predicted locations. **(c)** An AlphaFold2 predicted structure of human HAS2 (coloured blue to red from its N- to C-terminus) overlaid with the Cv-HAS structure shown as a grey cartoon and semi-transparent surface. **(d)** TMH1 remains disordered when two cytosolic nanobodies are used for cryo-EM analyses. **(e)** Catalytic activity of TMH1 truncated Cv-HAS. Left: Western blot of IMVs used for *in vitro* activity measurements. Right: Catalytic activity of the indicated Cv-HAS mutants expressed relative to the wild type enzyme. The assay quantifies  $^3\text{H}$ -labelled HA by scintillation counting. Control reactions in the absence of UDP-GlcA served as background and are subtracted. Error bars represent deviations from the means with  $n = 3$  independent experiments.

## Extended Data Fig. 6 Lipids plug the lateral channel opening.

**(a)** Representative map regions for modelled lipids contoured at  $7.0\sigma$  (from the UDP-GlcNAc bound set). **(b)** 2D slice from MD simulations of Cv-HAS (black area) within a POPE bilayer. Water and lipid densities are coloured blue and green, respectively. Right panel: Lipid contact times with selected channel residues. **(c)** Comparison of the Cv-HAS (rainbow coloured from the N- to C-terminus) and RsBcsA (grey, 4P00). Cellulose associated with BcsA is shown as black sticks. Helices are shown as cylinders except BcsA's N-terminal two TMHs, which are shown as coils.

## Extended Data Fig. 7 Details of substrate-binding and of priming-induced conformational changes.

**(a and b)** Map quality for UDP-GlcNAc, UDP, and  $\text{Mn}^{2+}$  ligands. **(c)** Comparison of UDP and UDP-GlcNAc positions. **(d)** Map for the priming loop in nucleotide bound states. **(e)** Representative map for the GlcNAc

primer. (f) Map for the priming loop in the primed states. (g) Contact point of TMH2 (open in blue, closed in grey) with IF1 in the primed state. (h and i) Map quality for TMH2 in a closed position (UDP-GlcNAc bound) and open position. All maps are contoured at  $7.0\sigma$ .

### Extended Data Fig. 8 GlcNAc priming of HA biosynthesis.

Shown is an autoradiogram of  $^{14}\text{C}$ -labelled HA after SDS-PAGE. The experiment has been repeated at least 3 times with essentially identical results.

### Extended Data Fig. 9 Effect of monosaccharides on substrate hydrolysis.

(a and b) Reaction schemes for UDP-GlcA and UDP-GlcNAc hydrolysis. (c and d) Raw absorbance measurements. (e) Quantification of hydrolysis rates in the presence of increasing monosaccharide concentrations. Blue and Red: Hydrolysis of UDP-GlcNAc and UDP-GlcA, respectively. Light and dark colours represent control reactions in the absence of enzyme. Right panel: Background subtracted hydrolysis rates. Error bars represent deviations from the means with  $n = 3$  independent experiments.

### Extended Data Fig. 10 Likely mechanism of alternating substrate polymerization and comparison with cellulose synthase.

(a) Superimposition of substrate-bound and primed Cv-HAS structures. The close distance between the primer and donor sugar is indicated by grey bars. (b) Contact likelihood between C231 and GlcNAc for the systems in a over the last 125 ns of each simulation. In the case of GlcNAc being in both donor and acceptor positions, both GlcNAc units are less likely to bind C231, and exhibit very high variance regarding binding poses. Of particular note, the chance of both GlcNAc being in the C231 pocket at the same time is very low (ca.  $1.5 \pm 0.8\%$ ). (c) Cv-HAS is superimposed with the *Rhodobacter sphaeroides* (Rs) BcsA-B complex (PDB: 4P00) based on secondary structure matching. Rs-BcsA-B is coloured grey and Cv-HAS is

coloured blue and green for its TM and GT domains. The cellulose polymer associated with Rs-BcsA-B is shown as black sticks.

## **Extended Data Table 1 Data collection, processing, and refinement statistics**

## **Supplementary information**

### **Supplementary Information**

Supplementary Discussions 1–3, Supplementary Tables 1–3 and Supplementary References.

### **Reporting Summary**

### **Peer Review File**

### **Supplementary Figure 1.**

### **Supplementary Video 1**

A continuous volume series representing density differences at the active site. Overlay of densities from a simple 3D variability display job corresponding to Fig. 2a. Trp342 is shown in sticks, UDP-GlcNAc and the GlcNAc primer are shown as ball and sticks coloured grey and cyan for carbon atoms, respectively.

### **Supplementary Video 2**

Movement of the priming loop. Output from a simple 3D variability display overlayed with a Chimera morph created from models of the Primed open and the UDP-GlcNAc-bound states. The video oscillates from the primed open state through the UDP-GlcNAc-bound state and back to the primed open state.

### **Supplementary Video 3**

Substrate-induced global conformational changes of Cv-HAS. Output from a simple 3D variability display overlayed with a Chimera morph created from models of the Primed open and the UDP-GlcNAc bound states. Oscillation from the primed open state through the UDP-GlcNAc-bound state and back to the primed open state.

## Supplementary Video 4

Tilting of the TMH2 and opening of the putative HA channel. Output from a simple 3D variability display overlayed with a Chimera morph created from models of the Primed open and the primed closed states. Blue, TMH2; orange, TMH4; green, TMH6. Oscillation from the primed closed state through the primed open state and back to the primed closed state.

## Rights and permissions

### Reprints and Permissions

## About this article

### Cite this article

Maloney, F.P., Kuklewicz, J., Corey, R.A. *et al.* Structure, substrate recognition and initiation of hyaluronan synthase. *Nature* **604**, 195–201 (2022). <https://doi.org/10.1038/s41586-022-04534-2>

- Received: 08 August 2021
- Accepted: 08 February 2022
- Published: 30 March 2022
- Issue Date: 07 April 2022
- DOI: <https://doi.org/10.1038/s41586-022-04534-2>

## Share this article

Anyone you share the following link with will be able to read this content:

[Get shareable link](#)

Sorry, a shareable link is not currently available for this article.

[Copy to clipboard](#)

Provided by the Springer Nature SharedIt content-sharing initiative

---

This article was downloaded by **calibre** from <https://www.nature.com/articles/s41586-022-04534-2>

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

[Download PDF](#)

- Matters Arising
- [Published: 06 April 2022](#)

# Evidence from a statewide vaccination RCT shows the limits of nudges

- [Nathaniel Rabb](#) ORCID: [orcid.org/0000-0002-7283-2549](https://orcid.org/0000-0002-7283-2549)<sup>1</sup>,
- [Megan Swindal](#)<sup>2</sup>,
- [David Glick](#)<sup>1,3</sup>,
- [Jake Bowers](#) ORCID: [orcid.org/0000-0002-4048-1166](https://orcid.org/0000-0002-4048-1166)<sup>1,4</sup>,
- [Anna Tomasulo](#)<sup>2</sup>,
- [Zayid Oyelami](#)<sup>5</sup>,
- [Kevin H. Wilson](#)<sup>1</sup> &
- [David Yokum](#)<sup>1</sup>

[Nature](#) volume 604, pages E1–E7 (2022)

- 1335 Accesses
- 50 Altmetric
- [Metrics details](#)

## Subjects

- [Epidemiology](#)
- [Human behaviour](#)

The [Original Article](#) was published on 02 August 2021

arising from H. Dai et al. *Nature* <https://doi.org/10.1038/s41586-021-03843-2> (2021)

Simple messages derived from behavioural science have increased the uptake of the seasonal flu vaccine<sup>1,2,3,4,5</sup>, and early studies from the coronavirus disease 2019 (COVID-19) vaccine rollout have found that this strategy works for recently eligible

older adults<sup>6</sup> and healthcare workers<sup>7</sup>. However, it is unknown whether messaging on its own will encourage vaccination against COVID-19 among reluctant populations. In a randomized controlled trial (RCT) five to eight weeks after all adults in the study population ( $n = 142,428$ ) were eligible for vaccination, we find that the best-performing nudge in previous studies<sup>2,6</sup> and seven additional messages—stressing vaccines’ safety, efficacy, minimization of bad outcomes, accessibility (free, no identification required), protection of recipients’ families or widespread adoption—had no detectable effect among people who had not been vaccinated according to state records. This suggests an important boundary condition for nudges that is consistent with a recent result from late in the flu season<sup>8</sup>. Public health authorities should consider simple messages to encourage vaccination at key inflection points (for example, rollout of paediatric COVID-19 vaccines and full Food and Drug Administration approval for adults), but may see diminishing returns if using them to encourage the more hesitant.

After a strong initial push, the rate of COVID-19 vaccinations declined in the USA. Efforts to encourage vaccination have run the gamut from free doughnuts and marijuana to million-dollar lotteries and rare experiences such as driving at a superspeedway. Recently, Dai et al.<sup>6</sup> reported promising results from an RCT evaluating another tactic—sending people short messages informed by behavioural science. The appeal of this approach is clear: it is cheap and minimally invasive. It is also well supported by convergent evidence: email messages increased COVID-19 vaccination appointment sign-ups among healthcare workers<sup>7</sup>, and SMS<sup>1,2,3</sup>, mail<sup>4</sup> and email<sup>5</sup> messages have increased seasonal flu vaccinations. Moreover, it has garnered considerable media attention<sup>9</sup>, with pieces advocating it in *The Washington Post*, *Fortune*, *The Guardian*, *U.S. News & World Report* and this journal<sup>10</sup>. Policymakers also took note, as several states implemented SMS campaigns<sup>9</sup>.

The Dai et al. study was conducted early in the COVID-19 vaccine rollout with recently eligible older adults. Although the results show the potential of nudges, it is unknown whether short messages can change motivations in the population that did not get vaccinated immediately. Indeed, Dai et. al. distinguish burden reduction (helping people to follow through on pre-existing intentions) from demand creation (changing intentions), and numerous reviews find limited and mixed evidence on what drives demand<sup>11,12,13,14</sup>.

To test whether these findings generalize beyond the initial stages of COVID-19 vaccination, we evaluated the efficacy of text messages sent by the Rhode Island Department of Health (RIDOH) to increase uptake in May and June 2021. The messages included the best-performing ‘ownership’ language from Dai et al. and a related flu study<sup>2</sup>. This language was supplemented in most conditions with information about safety, efficacy or access, for example. This study offers a strong

test of direct messaging because recipients were unvaccinated five to eight weeks after becoming eligible. It is also a realistic test of what a government can and, more importantly, cannot do (for example, craft messages containing false claims and send excessive communications).

RIDOH maintains separate databases of individuals who have been vaccinated and tested for COVID-19. Our study population is the difference of these lists (tested but not yet vaccinated) matched through a series of quasi-identifiers and excluding people under 18 when tested (final  $n = 142,428$ ; see Extended Data Fig. 1 for randomization scheme). The primary outcome was vaccination by the end of the measurement period: 25 May 2021 to 21 June 2021 (one week after the last day of messaging). At time of launch, all Rhode Islanders over 16 had been eligible to get vaccinated since 19 April 2021, and free, walk-in availability was widespread. The study was deemed exempt by RIDOH's institutional review board. The sample size was dictated by policy goals, as all eligible individuals received messages. A previous study<sup>2</sup> with more conditions and a sample size similar to our first iteration detected meaningful effects.

We created eight messages (Extended Data Table 1, Supplementary Information section 1) on the basis of behavioural science research on COVID-19 health behaviours and other vaccination contexts. All included ownership language ('a vaccine is waiting for you')<sup>2,6</sup>, a sentiment also appearing in a standalone condition. Other conditions further emphasized safety, access, minimal likelihood of bad outcomes, reduced risk to one's family, social norms or some combination. All included a link to a state-run page providing vaccination options.

Individuals were assigned to receive one of eight messages or no message (control group). We randomly divided the population into three consecutive iterations of 40,000, 39,709 or 78,394, and then into roughly equal groups per day within those weeks. Within these strata, individuals were assigned to receive one of eight messages or no message (control group).

To maximize overall vaccinations, in iterations 2 and 3 we used an adaptive design such that the likelihood of assignment to any given message was determined by message performance in the previous iteration, with an  $\varepsilon$ -bounded Thompson sampler adjusting the probability of assignment to condition over time (Supplementary Information section 2).

This study is a block-randomized experiment. All analyses (pre-registration: <https://osf.io/pkhae>) use either the Cochran–Mantel–Haenszel (CMH) test for 9 (condition)  $\times$  2 (outcome)  $\times$  13 (day) strata tables or a block-specific weighting, which provides unbiased estimates of intent-to-treat effects and randomization-justified variance calculations.

No SMS message did substantially better or worse than the control whether vaccination rates were measured one week after the messages were sent or at the end of the study period. Figure 1 illustrates the small size of these differences: the largest positive difference was 0.002 for the ‘preventing bad outcomes’ condition (that is, 2% of control and 2.2% of ‘preventing bad outcomes’ were vaccinated). Furthermore, we see no evidence of differences in vaccination rates (however measured) between the control and an aggregated ‘any message’ condition (estimated difference in proportions vaccinated  $-0.001$ , 95% confidence interval (CI)  $-0.004$  to  $0.001$ , CMH test,  $P = 0.27$ ), nor between the arms taken all together (CMH test for  $9 \times 2 \times 13$  table,  $P = 0.12$ ). For demographics, see Extended Data Table 2; for additional analyses see Supplementary Information sections 3–6.

**Fig. 1: Average treatment effects for the eight experimental conditions overall and proportions vaccinated by day.**

 figure 1

Top left, the differences in the proportion vaccinated by the end of the study between each message condition and the control or ‘no message’ condition (2% of the control condition was vaccinated within the study period). Top right, the differences in the proportion vaccinated within a week of message sending (1% of the control condition

was vaccinated within a week of message sending). The total control condition participation was 11,327. The total size of each arm is shown on the right. All point estimates with 95% confidence intervals (CIs). No adjustment was made for multiple testing as no test cast doubt on the null of no difference. Bottom, proportions vaccinated by 22 June 2021 in each message by the date messages were sent. The grey vertical line shows the proportion vaccinated in the control condition. The 95% confidence intervals for small proportions come from the binomial ensemble method of ref. [17](#).

We find no evidence that a strategy found effective early in the vaccine rollout<sup>[6,7](#)</sup> increased COVID-19 vaccination among people who remained unvaccinated five or more weeks after becoming eligible. Public health officials—especially those avoiding or legally barred from mandates—may turn to this strategy to increase vaccination rates among the less enthusiastic but will probably see minimal impact. Dai et al. highlighted a promising, valuable and low-cost tool that can help to increase vaccinations; although our result does not contradict theirs, it does bound the reach of such approaches, a possibility one of their co-authors contemplated elsewhere<sup>[10](#)</sup>.

One limitation of our study is that the initial recipient list may contain some vaccinated people. Rhode Island residents could get tested at home but vaccinated out of state, and certain sites (for example, Veterans Affairs hospitals) do not need to report individual-level records to the state. Base rates may be inaccurate because of this and other sources of noise (Supplementary Information section [6](#)), although this would not mask treatment effects, as message assignment was random. Another limitation is that race and ethnicity information is incomplete (Extended Data Table [2](#)).

The study by Dai et al. differed from ours in several ways, including population age (mean age 70 versus 39), message source (recipients' health network versus a state agency), sign-up ease (recipients being directed to a sign-up system versus a page providing vaccination options) and vaccination context (appointments were scarce in February 2021 but abundant by May 2021). Although these factors could account for the different outcomes, flu vaccine findings suggest otherwise: similar interventions have shown success among younger populations<sup>[1](#)</sup>, when issued by the state<sup>[15](#)</sup>, and using inconvenient media (mailed letters<sup>[4](#)</sup>), and flu vaccines are comparatively easy to procure. One feature that Dai et al. and many flu vaccine studies do share is that they were conducted early in their respective campaigns, whereas ours was not. Notably, a study of older adults found increased uptake of flu vaccines due to postcard messages in October but not November, December or January<sup>[8](#)</sup>. Taken together, this suggests that nudges help early in vaccination campaigns, but the efficacy decays. Another COVID-19 study recently made public provides further support<sup>[16](#)</sup>.

Although we cannot identify the mechanism(s) responsible for decaying efficacy of nudges, the possibilities include novelty effects early on, oversaturation effects later on, different types of hesitancy (logistical barriers versus objections to vaccines), and, especially for COVID-19, increasingly polarized discourse, divergent social norms and differential vaccine knowledge. Future work in public health communication should distinguish these mechanisms to better implement message campaigns. It may also be that short messages effectively encourage those somewhat inclined to vaccinate but cannot move those less inclined, regardless of timing, and with time, the former group shrinks. Despite our null result, nudges may serve foreseeable public health needs (for example, vaccinating children under 5 or promoting boosters) if timed correctly. Indeed, we know of no studies showing reduced vaccinations owing to message campaigns, so they carry little potential harm. However, their ability to move the more reluctant may be limited.

## Reporting summary

Further information on experimental design is available in the [Nature Research Reporting Summary linked](#) to this paper.

## Data availability

The data analysed in this paper were provided by the Rhode Island Department of Health and contains protected health information. To protect privacy, we cannot publicly post individual-level data. Qualified researchers with a valuable research question and relevant approvals including ethical approval can request access to the de-identified data about this trial from the corresponding author. A formal contract will be signed and an independent data protection agency should oversee the sharing process to ensure the safety of the data. Lightly aggregated data that support most of the analyses in this paper can be found at <https://github.com/thepolicylab/COVID-SMSExperiment>. Some demographic analyses rely on publicly available data from the United States Census Bureau, the United States Department of Housing and Urban Development, the Rhode Island Geographical Information System and the Rhode Island Board of Elections. Copies of these data and, where appropriate, the code that gathered the data are available at <https://github.com/thepolicylab/COVID-SMSExperiment>.

## Code availability

The code to replicate the analyses and figures in the paper and the Extended Data is available at <https://github.com/thepolicylab/COVID-SMSExperiment>.

## References

1. Lee, W.-N. et al. Large-scale influenza vaccination promotion on a mobile app platform: a randomized controlled trial. *Vaccine* **38**, 3508–3514 (2020).
2. Milkman, K. L. et al. A megastudy of text-based nudges encouraging patients to get vaccinated at an upcoming doctor’s appointment. *Proc. Natl Acad. Sci. USA* **118**, e2101165118 (2021).
3. Regan, A. K., Bloomfield, L., Peters, I. & Effler, P. V. Randomized controlled trial of text message reminders for increasing influenza vaccination. *Ann. Fam. Med.* **15**, 507–514 (2017).
4. Yokum, D., Lauffenburger, J. C., Ghazinouri, R. & Choudhry, N. K. Letters designed with behavioural science increase influenza vaccination in Medicare beneficiaries. *Nat. Hum. Behav.* **2**, 743–749 (2018).
5. Baskin, E. Increasing influenza vaccination rates via low cost messaging interventions. *PLoS ONE* **13**, e0192594 (2018).
6. Dai, H. et al. Behavioral nudges increase COVID-19 vaccinations. *Nature* **597**, 404–409 (2021).
7. Santos, H. C., Goren, A., Chabris, C. F. & Meyer, M. N. Effect of targeted behavioral science messages on COVID-19 vaccination registration among employees of a large health system: a randomized trial. *JAMA Netw. Open* **4**, e2118702 (2021).
8. Chen, N., Trump, K.-S., Hall S. & Le Q. The effect of postcard reminders on vaccinations among the elderly: a block-randomized experiment. *Behav. Public Policy* <https://doi.org/10.1017/bpp.2020.34> (2020).
9. Mascarenhas, L. This simple text message can encourage people to get vaccinated, researchers say. *CNN* <https://www.cnn.com/2021/06/29/health/text-messages-vaccination-wellness/index.html> (2021).
10. Kozlov, M. A simple text has the power to increase COVID vaccinations. *Nature* <https://doi.org/10.1038/d41586-021-02108-2> (2021).
11. Brewer, N. T., Chapman, G. B., Rothman, A. J., Leask, J. & Kempe, A. Increasing vaccination: putting psychological science into action. *Psychol. Sci. Public Interest* **18**, 149–207 (2017).

12. Dubé, E., Gagnon, D. & MacDonald, N. E. Strategies intended to address vaccine hesitancy: review of published reviews. *Vaccine* **33**, 4191–4203 (2015).
13. Jarrett, C., Wilson, R., O’Leary, M., Eckersberger, E. & Larson, H. J. Strategies for addressing vaccine hesitancy: a systematic review. *Vaccine* **33**, 4180–4190 (2015).
14. Lawes-Wickwar, S. et al. A rapid systematic review of public responses to health messages encouraging vaccination against infectious diseases in a pandemic or epidemic. *Vaccines* **9**, 72 (2021).
15. Esteban-Vasallo, M. et al. Effect of mobile phone text messaging for improving the uptake of influenza vaccination in patients with rare diseases. *Vaccine* **37**, 5257–5264 (2019).
16. Chang, T. et al. Financial incentives and other nudges do not increase COVID-19 vaccinations among the vaccine hesitant. *NBER* <https://doi.org/10.3386/w29403> (2021).
17. Park, H. & Leemis, L. M. Ensemble confidence intervals for binomial proportions. *Stat. Med.* **38**, 3460–3475 (2019).
18. Milkman, K. L. et al. A mega-study of text-message nudges encouraging patients to get vaccinated at their pharmacy. *SSRN* <https://doi.org/10.2139/ssrn.3780356> (2021).
19. Hamel, L. et al. KFF COVID-19 vaccine monitor: April 2021. *KFF* <http://www.kff.org/coronavirus-covid-19/poll-finding/kff-covid-19-vaccine-monitor-april-2021> (2021).
20. Callaghan, T. et al. Correlates and disparities of COVID-19 vaccine hesitancy. *SSRN* <https://doi.org/10.2139/ssrn.3667971> (2020).
21. Gibson, D. et al. COVID-19 vaccine acceptability and inequity in the United States: results from a nationally representative survey. Preprint at *MedRxiv* <https://doi.org/10.1101/2021.01.29.21250784> (2021).
22. Huynh, H. Most (but not all) anti-vaccination attitudes increased alongside the occurrence and persistence of COVID-19: a snapshot of the trends in the U.S. Preprint at *PsyArXiv* <https://doi.org/10.31234/osf.io/yzu9r> (2020).
23. Kaplan, R. M. & Milstein, A. Influence of a COVID-19 vaccine’s effectiveness and safety profile on vaccination acceptance. *Proc. Natl Acad. Sci. USA* **118**, e2021726118 (2021).

24. Reiter, P. L., Pennell, M. L. & Katz, M. L. Acceptability of a COVID-19 vaccine among adults in the United States: how many people would get vaccinated? *Vaccine* **38**, 6500–6507 (2020).
25. Taylor, S. et al. A proactive approach for managing COVID-19: the importance of understanding the motivational roots of vaccination hesitancy for SARS-CoV2. *Front. Psychol.* **11**, 575950 (2020).
26. Karafillakis, E. & Larson, H. J. The benefit of the doubt or doubts over benefits? A systematic literature review of perceived risks of vaccines in European populations. *Vaccine* **35**, 4840–4850 (2017).
27. Yaqub, O. et al. Attitudes to vaccination: a critical review. *Soc. Sci. Med.* **112**, 1–11 (2014).
28. Hallsworth, M., Mirpuri, S. & Toth, C. Four messages that can increase uptake of the COVID-19 vaccines. *Behavioural Insights Team*   
<https://www.bi.team/blogs/four-messages-that-can-increase-uptake-of-the-covid-19-vaccines/> (2021).
29. Hamel, L. et al. KFF COVID-19 vaccine monitor: March 2021. *KFF*   
<https://www.kff.org/coronavirus-covid-19/poll-finding/kff-covid-19-vaccine-monitor-march-2021/> (2021).
30. Prochaska, J. O. & Velicer, W. F. The transtheoretical model of health behavior change. *Am. J. Health Promot.* **12**, 38–48 (1997).
31. Amin, A. B. et al. Association of moral values with vaccine hesitancy. *Nat. Hum. Behav* **1**, 873–880 (2017).
32. de Figueiredo, A. et al. Mapping global trends in vaccine confidence and investigating barriers to vaccine uptake: a large-scale retrospective temporal modelling study. *Lancet* **396**, 898–908 (2020).
33. Kerr, J. R. et al. Correlates of intended COVID-19 vaccine acceptance across time and countries: results from a series of cross-sectional surveys. *BMJ Open* **11**, e048025 (2021).
34. Larson, H. J. et al. Measuring trust in vaccination: a systematic review. *Hum. Vac. Immunother.* **14**, 1599–1609 (2017).
35. Linholdt, M. F. et al. Public acceptance of COVID-19 vaccines: cross-national evidence on levels and individual-level predictors using observational data. *BMJ Open* **11**, e048172 (2021).

36. Mesch, G. S. & Schwirian, K. P. Social and political determinants of vaccine hesitancy: lessons learned from the H1N1 pandemic of 2009–2010. *Am. J. Infect. Control* **43**, 1161–1165 (2015).
37. Bunker, S. & Park, J. Evaluating prosocial COVID-19 messaging frames: evidence from a field study on Facebook. *Judgm. Decis. Mak.* **15**, 1037–1043 (2020).
38. Rabb, N. et al. No evidence that collective-good appeals best promote COVID-related health behaviors. *Proc. Natl Acad. Sci. USA* **118**, e2100662118 (2021).
39. Falco, P. & Zaccagni, S. Promoting social distancing in a pandemic: beyond good intentions. *PLoS ONE* **16**, e0260457 (2021).
40. Bicchieri, C. et al. In science we (should) trust: expectations and compliance across nine countries during the COVID-19 pandemic. *PLoS ONE* **16**, e0252892 (2021).
41. Goldberg, M. H. et al. Social norms motivate COVID-19 preventive behaviors. Preprint at *PsyArXiv* <https://doi.org/10.31234/osf.io/9whp4> (2020).
42. Lees, J. et al. Intentions to comply with COVID-19 preventive behaviors are associated with personal beliefs, independent of perceived social norms. Preprint at *PsyArXiv* <https://doi.org/10.31234/osf.io/97jry> (2020).
43. Bruine de Bruin, W. et al. Reports of social circles' and own vaccination behavior: a national longitudinal survey. *Health Psychol.* **38**, 975–983 (2019).
44. Kumar, S. et al. The social ecological model as a framework for determinants of 2009 H1N1 influenza vaccine uptake in the United States. *Health Educ. Behav.* **39**, 229–243 (2012).
45. Sheeran, P. et al. The impact of changing attitudes, norms, and self-efficacy on health-related intentions and behavior: a meta-analysis. *Health Psychol.* **35**, 1178–1188 (2016).
46. Sinclair, S. & Agerström, J. Do social norms influence young people's willingness to take the COVID-19 vaccine? *Health Commun.* <https://doi.org/10.1080/10410236.2021.1937832> (2021).
47. Moehring, A. et al. Surfacing norms to increase vaccine acceptance. Preprint at *PsyArXiv* <https://doi.org/10.31234/osf.io/srv6t> (2021).

48. Schmidtke, K. A. et al. Randomised controlled trial of a theory-based intervention to prompt front-line staff to take up the seasonal influenza vaccine. *BMJ Qual. Saf.* **29**, 189–197 (2020).

## Author information

### Affiliations

1. The Policy Lab at Brown University, Providence, RI, USA  
Nathaniel Rabb, David Glick, Jake Bowers, Kevin H. Wilson & David Yokum
2. Rhode Island Department of Health, Providence, RI, USA  
Megan Swindal & Anna Tomasulo
3. Department of Political Science, Boston University, Boston, MA, USA  
David Glick
4. Department of Political Science, University of Illinois, Urbana-Champaign, IL, USA  
Jake Bowers
5. IBM, Armonk, NY, USA  
Zayid Oyelami

### Contributions

N.R., M.S., D.G., J.B., A.T., K.H.W. and D.Y. conceived of and designed the study. Z.O. and K.H.W. oversaw data collection. J.B., K.H.W., D.G. and N.R. conducted the analysis. J.B., K.H.W., D.G., N.R. and M.S. interpreted the data. All authors contributed to the manuscript.

### Corresponding author

Correspondence to [Nathaniel Rabb](#).

### Ethics declarations

## Competing interests

The authors declare no competing interests.

## Additional information

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## Extended data figures and tables

### Extended Data Fig. 1 Randomization scheme and sample.

RIDOH maintains separate databases of (a) individuals who have been vaccinated and (b) individuals who have been tested for COVID-19. Vaccination data comes from medical providers and pharmacies receiving vaccines supplied by the State of Rhode Island, who are required to participate in the Rhode Island Child and Adult Immunization Registry (RICAIR) through electronic data reporting. Immunization records can be accessed by an individual's medical provider or by authorized RIDOH users conducting public health surveillance activities including linking vaccination records with the state's COVID-19 testing or case databases to verify information collected during case investigation. COVID-19 testing data (b) is reported to the state through the National Electronic Disease Surveillance System (NEDSS). Our study population is the difference of lists (a) and (b); the resulting database contained 162,504 unique entries. The study ended one day early after RIDOH received complaints about excessive communication. It is unclear how many complaints were received and how many were specifically about this study; other concurrent outreach efforts included SMS messages about COVID-19 testing and phone calls to older adults encouraging vaccination. Nevertheless, leadership halted all such communications out of concern that people would block crucial emergency messages. The final  $N$  for the study is 142,428. A small subset of the initial population ( $N \approx 800$ ) had chosen Spanish as their preferred language on testing sign-up forms. While we had initially planned to send this group messages translated into Spanish, an unresolved encoding problem prevented Spanish characters from displaying properly on some cell phones. The project team decided to reintroduce these individuals into the general study population for Iteration 3.

**Extended Data Table 1 Messages used in the RCT and rationales**

**Extended Data Table 2 Demographics for study population**

## Supplementary information

## Supplementary Information

This file contains Supplementary Information, including Supplementary Figs. 1, 2 and additional references.

## Reporting Summary

## Rights and permissions

### Reprints and Permissions

## About this article

### Cite this article

Rabb, N., Swindal, M., Glick, D. *et al.* Evidence from a statewide vaccination RCT shows the limits of nudges. *Nature* **604**, E1–E7 (2022).  
<https://doi.org/10.1038/s41586-022-04526-2>

- Received: 18 August 2021
- Accepted: 01 February 2022
- Published: 06 April 2022
- Issue Date: 07 April 2022
- DOI: <https://doi.org/10.1038/s41586-022-04526-2>

## Share this article

Anyone you share the following link with will be able to read this content:

[Get shareable link](#)

Sorry, a shareable link is not currently available for this article.

[Copy to clipboard](#)

Provided by the Springer Nature SharedIt content-sharing initiative

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

# Amendments & Corrections

- [\*\*Author Correction: Comparative cellular analysis of motor cortex in human, marmoset and mouse\*\*](#) [ 22 March 2022]  
Author Correction •
- [\*\*Publisher Correction: Differential mechanisms underlie trace and delay conditioning in Drosophila\*\*](#) [ 22 March 2022]  
Publisher Correction •
- [\*\*Publisher Correction: Structural basis for mismatch surveillance by CRISPR–Cas9\*\*](#) [ 22 March 2022]  
Publisher Correction •
- [\*\*Publisher Correction: ER proteins decipher the tubulin code to regulate organelle distribution\*\*](#) [ 23 March 2022]  
Publisher Correction •
- [\*\*Retraction Note: A 10 per cent increase in global land evapotranspiration from 2003 to 2019\*\*](#) [ 24 February 2022]  
Retraction •

[Download PDF](#)

- Author Correction
- Open Access
- [Published: 22 March 2022](#)

# Author Correction: Comparative cellular analysis of motor cortex in human, marmoset and mouse

- [Trygve E. Bakken](#) [ORCID: orcid.org/0000-0003-3373-7386<sup>1</sup>](#),
- [Nikolas L. Jorstad<sup>1</sup>](#),
- [Qiwen Hu<sup>2</sup>](#),
- [Blue B. Lake<sup>3</sup>](#),
- [Wei Tian<sup>4</sup>](#),
- [Brian E. Kalmbach](#) [ORCID: orcid.org/0000-0003-3136-8097<sup>1,5</sup>](#),
- [Megan Crow](#) [ORCID: orcid.org/0000-0002-1172-5897<sup>6</sup>](#),
- [Rebecca D. Hodge](#) [ORCID: orcid.org/0000-0002-5784-9668<sup>1</sup>](#),
- [Fenna M. Krienen](#) [ORCID: orcid.org/0000-0002-1400-6820<sup>7</sup>](#),
- [Staci A. Sorensen<sup>1</sup>](#),
- [Jeroen Eggermont<sup>8</sup>](#),
- [Zizhen Yao](#) [ORCID: orcid.org/0000-0002-9361-5607<sup>1</sup>](#),
- [Brian D. Aevermann<sup>9</sup>](#),
- [Andrew I. Aldridge](#) [ORCID: orcid.org/0000-0003-1962-8802<sup>10</sup>](#),
- [Anna Bartlett<sup>10</sup>](#),
- [Darren Bertagnolli<sup>1</sup>](#),
- [Tamara Casper<sup>1</sup>](#),
- [Rosa G. Castanon<sup>10</sup>](#),
- [Kirsten Crichton<sup>1</sup>](#),
- [Tanya L. Daigle<sup>1</sup>](#),

- [Rachel Dalley](#)<sup>1</sup>,
- [Nick Dee](#)<sup>1</sup>,
- [Nikolai Dembrow](#) ORCID: [orcid.org/0000-0002-7699-6591](#)<sup>5,11</sup>,
- [Dinh Diep](#)<sup>3</sup>,
- [Song-Lin Ding](#)<sup>1</sup>,
- [Weixiu Dong](#)<sup>3</sup>,
- [Rongxin Fang](#)<sup>12</sup>,
- [Stephan Fischer](#)<sup>6</sup>,
- [Melissa Goldman](#)<sup>7</sup>,
- [Jeff Goldy](#)<sup>1</sup>,
- [Lucas T. Graybuck](#) ORCID: [orcid.org/0000-0002-8814-6818](#)<sup>1</sup>,
- [Brian R. Herb](#)<sup>13</sup>,
- [Xiaomeng Hou](#)<sup>14</sup>,
- [Jayaram Kancherla](#)<sup>15</sup>,
- [Matthew Kroll](#)<sup>1</sup>,
- [Kanan Lathia](#)<sup>1</sup>,
- [Baldur van Lew](#) ORCID: [orcid.org/0000-0003-0628-1264](#)<sup>8</sup>,
- [Yang Eric Li](#) ORCID: [orcid.org/0000-0001-6997-6018](#)<sup>14,16</sup>,
- [Christine S. Liu](#)<sup>17,18</sup>,
- [Hanqing Liu](#)<sup>10</sup>,
- [Jacinta D. Lucero](#)<sup>4</sup>,
- [Anup Mahurkar](#) ORCID: [orcid.org/0000-0002-4999-2296](#)<sup>13</sup>,
- [Delissa McMillen](#)<sup>1</sup>,
- [Jeremy A. Miller](#) ORCID: [orcid.org/0000-0003-4549-588X](#)<sup>1</sup>,
- [Marmar Moussa](#)<sup>19</sup>,
- [Joseph R. Nery](#) ORCID: [orcid.org/0000-0003-0153-5659](#)<sup>10</sup>,
- [Philip R. Nicovich](#)<sup>1</sup>,
- [Sheng-Yong Niu](#)<sup>10,20</sup>,
- [Joshua Orvis](#)<sup>13</sup>,
- [Julia K. Osteen](#) ORCID: [orcid.org/0000-0001-7058-3297](#)<sup>4</sup>,
- [Scott Owen](#)<sup>1</sup>,
- [Carter R. Palmer](#)<sup>17,18</sup>,
- [Thanh Pham](#) ORCID: [orcid.org/0000-0002-4738-5062](#)<sup>1</sup>,
- [Nongluk Plongthongkum](#)<sup>3</sup>,

- [Olivier Poirion](#)<sup>14</sup>,
- [Nora M. Reed](#)<sup>7</sup>,
- [Christine Rimorin](#)<sup>1</sup>,
- [Angeline Rivkin](#)<sup>4</sup>,
- [William J. Romanow](#)<sup>17</sup>,
- [Adriana E. Sedeño-Cortés](#)<sup>1</sup>,
- [Kimberly Siletti](#)<sup>21</sup>,
- [Saroja Somasundaram](#)<sup>1</sup>,
- [Josef Sulc](#)<sup>1</sup>,
- [Michael Tieu](#)<sup>1</sup>,
- [Amy Torkelson](#)<sup>1</sup>,
- [Herman Tung](#) ORCID: [orcid.org/0000-0002-0812-3318](http://orcid.org/0000-0002-0812-3318)<sup>1</sup>,
- [Xinxin Wang](#) ORCID: [orcid.org/0000-0001-6393-2276](http://orcid.org/0000-0001-6393-2276)<sup>22</sup>,
- [Fangming Xie](#)<sup>23</sup>,
- [Anna Marie Yanny](#)<sup>1</sup>,
- [Renee Zhang](#)<sup>9</sup>,
- [Seth A. Ament](#) ORCID: [orcid.org/0000-0001-6443-7509](http://orcid.org/0000-0001-6443-7509)<sup>13</sup>,
- [M. Margarita Behrens](#) ORCID: [orcid.org/0000-0002-7168-8186](http://orcid.org/0000-0002-7168-8186)<sup>4</sup>,
- [Hector Corrada Bravo](#)<sup>15</sup>,
- [Jerold Chun](#) ORCID: [orcid.org/0000-0003-3964-0921](http://orcid.org/0000-0003-3964-0921)<sup>17</sup>,
- [Alexander Dobin](#)<sup>24</sup>,
- [Jesse Gillis](#)<sup>6</sup>,
- [Ronna Hertzano](#)<sup>25</sup>,
- [Patrick R. Hof](#)<sup>26</sup>,
- [Thomas Höllt](#) ORCID: [orcid.org/0000-0001-8125-1650](http://orcid.org/0000-0001-8125-1650)<sup>27</sup>,
- [Gregory D. Horwitz](#) ORCID: [orcid.org/0000-0001-5130-5259](http://orcid.org/0000-0001-5130-5259)<sup>28</sup>,
- [C. Dirk Keene](#) ORCID: [orcid.org/0000-0002-5291-1469](http://orcid.org/0000-0002-5291-1469)<sup>29</sup>,
- [Peter V. Kharchenko](#) ORCID: [orcid.org/0000-0002-6036-5875](http://orcid.org/0000-0002-6036-5875)<sup>2</sup>,
- [Andrew L. Ko](#)<sup>30,31</sup>,
- [Boudewijn P. Lelieveldt](#) ORCID: [orcid.org/0000-0001-8269-7603](http://orcid.org/0000-0001-8269-7603)<sup>8,32</sup>,
- [Chongyuan Luo](#)<sup>33</sup>,
- [Eran A. Mukamel](#) ORCID: [orcid.org/0000-0003-3203-9535](http://orcid.org/0000-0003-3203-9535)<sup>34</sup>,
- [António Pinto-Duarte](#) ORCID: [orcid.org/0000-0002-2215-7653](http://orcid.org/0000-0002-2215-7653)<sup>4</sup>,
- [Sebastian Preiss](#) ORCID: [orcid.org/0000-0001-8971-5616](http://orcid.org/0000-0001-8971-5616)<sup>14</sup>,

- [Aviv Regev](#) ORCID: orcid.org/0000-0003-3293-3158<sup>35</sup>,
- [Bing Ren](#) ORCID: orcid.org/0000-0002-5435-1127<sup>14,16</sup>,
- [Richard H. Scheuermann](#) ORCID: orcid.org/0000-0003-1355-892X<sup>9,36,37</sup>,
- [Kimberly Smith](#) ORCID: orcid.org/0000-0002-3142-1970<sup>1</sup>,
- [William J. Spain](#)<sup>5,11</sup>,
- [Owen R. White](#)<sup>13</sup>,
- [Christof Koch](#) ORCID: orcid.org/0000-0001-6482-8067<sup>1</sup>,
- [Michael Hawrylycz](#) ORCID: orcid.org/0000-0002-5741-8024<sup>1</sup>,
- [Basiljka Tasic](#) ORCID: orcid.org/0000-0002-6861-4506<sup>1</sup>,
- [Evan Z. Macosko](#) ORCID: orcid.org/0000-0002-2794-5165<sup>35</sup>,
- [Steven A. McCarroll](#) ORCID: orcid.org/0000-0002-6954-8184<sup>7,35</sup>,
- [Jonathan T. Ting](#)<sup>1,5</sup>,
- [Hongkui Zeng](#) ORCID: orcid.org/0000-0002-0326-5878<sup>1</sup>,
- [Kun Zhang](#) ORCID: orcid.org/0000-0002-7596-5224<sup>3</sup>,
- [Guoping Feng](#) ORCID: orcid.org/0000-0002-8021-277X<sup>38,39,40</sup>,
- [Joseph R. Ecker](#) ORCID: orcid.org/0000-0001-5799-5895<sup>10,41</sup>,
- [Sten Linnarsson](#) ORCID: orcid.org/0000-0002-3491-3444<sup>21</sup> &
- [Ed S. Lein](#) ORCID: orcid.org/0000-0001-9012-6552<sup>1</sup>

*Nature* volume 604, page E8 (2022)

- 476 Accesses
- 1 Altmetric
- [Metrics details](#)

## Subjects

- [Cellular neuroscience](#)
- [Genetics of the nervous system](#)
- [Molecular evolution](#)
- [Molecular neuroscience](#)

The [Original Article](#) was published on 06 October 2021

Correction to: *Nature* <https://doi.org/10.1038/s41586-021-03465-8>

Published online 6 October 2021

In the version of this article initially published, the Acknowledgements section was incomplete and has now been amended to include the following: “NIH BRAIN Initiative awards U01 MH121282 to J.R.E and M.M.B, U19 MH114831 to J.R.E. and E.M.C., U19 MH114830 to H.Z., U01 MH114819 to G.F., 1U01MH114828 to K.Z. and J.C., RF1MH123220 to M.H. and R.H.S., and U19 MH114821. NIH awards R01DC019370 to R.H., R24MH114815 to R.H. and O.R.W., and R24 MH114788 to O.R.W. Nancy and Buster Alvord Endowment to C.D.K.” The changes have been made to the HTML and PDF versions of the article

## Author information

### Affiliations

#### 1. Allen Institute for Brain Science, Seattle, WA, USA

Trygve E. Bakken, Nikolas L. Jorstad, Brian E. Kalmbach, Rebecca D. Hodge, Staci A. Sorensen, Zizhen Yao, Darren Bertagnolli, Tamara Casper, Kirsten Crichton, Tanya L. Daigle, Rachel Dalley, Nick Dee, Song-Lin Ding, Jeff Goldy, Lucas T. Graybuck, Matthew Kroll, Kanan Lathia, Delissa McMillen, Jeremy A. Miller, Philip R. Nicovich, Scott Owen, Thanh Pham, Christine Rimorin, Adriana E. Sedeño-Cortés, Saroja Somasundaram, Josef Sulc, Michael Tieu, Amy Torkelson, Herman Tung, Anna Marie Yanny, Kimberly Smith, Christof Koch, Michael Hawrylycz, Bosiljka Tasic, Jonathan T. Ting, Hongkui Zeng & Ed S. Lein

#### 2. Department of Biomedical Informatics, Harvard Medical School, Boston, MA, USA

Qiwen Hu & Peter V. Kharchenko

3. Department of Bioengineering, University of California, San Diego,  
La Jolla, CA, USA

Blue B. Lake, Dinh Diep, Weixiu Dong, Nongluk  
Plongthongkum & Kun Zhang

4. The Salk Institute for Biological Studies, La Jolla, CA, USA

Wei Tian, Jacinta D. Lucero, Julia K. Osteen, Angeline Rivkin, M.  
Margarita Behrens & António Pinto-Duarte

5. Department of Physiology and Biophysics, University of Washington,  
Seattle, WA, USA

Brian E. Kalmbach, Nikolai Dembrow, William J. Spain & Jonathan T.  
Ting

6. Stanley Institute for Cognitive Genomics, Cold Spring Harbor  
Laboratory, Cold Spring Harbor, NY, USA

Megan Crow, Stephan Fischer & Jesse Gillis

7. Department of Genetics, Harvard Medical School, Boston, MA, USA

Fenna M. Krienen, Melissa Goldman, Nora M. Reed & Steven A.  
McCarroll

8. LKEB, Department of Radiology, Leiden University Medical Center,  
Leiden, The Netherlands

Jeroen Eggermont, Baldur van Lew & Boudewijn P. Lelieveldt

9. J. Craig Venter Institute, La Jolla, CA, USA

Brian D. Aevermann, Renee Zhang & Richard H. Scheuermann

10. Genomic Analysis Laboratory, The Salk Institute for Biological  
Studies, La Jolla, CA, USA

Andrew I. Aldridge, Anna Bartlett, Rosa G. Castanon, Hanqing Liu, Joseph R. Nery, Sheng-Yong Niu & Joseph R. Ecker

11. Epilepsy Center of Excellence, Department of Veterans Affairs Medical Center, Seattle, WA, USA

Nikolai Dembrow & William J. Spain

12. Bioinformatics and Systems Biology Graduate Program, University of California, San Diego, La Jolla, CA, USA

Rongxin Fang

13. Institute for Genomes Sciences, University of Maryland School of Medicine, Baltimore, MD, USA

Brian R. Herb, Anup Mahurkar, Joshua Orvis, Seth A. Ament & Owen R. White

14. Center for Epigenomics, Department of Cellular and Molecular Medicine, University of California, San Diego, La Jolla, CA, USA

Xiaomeng Hou, Yang Eric Li, Olivier Poirion, Sebastian Preiss & Bing Ren

15. Department of Computer Science, University of Maryland College Park, College Park, MD, USA

Jayaram Kancherla & Hector Corrada Bravo

16. Ludwig Institute for Cancer Research, La Jolla, CA, USA

Yang Eric Li & Bing Ren

17. Sanford Burnham Prebys Medical Discovery Institute, La Jolla, CA, USA

Christine S. Liu, Carter R. Palmer, William J. Romanow & Jerold Chun

18. Biomedical Sciences Program, School of Medicine, University of California, San Diego, La Jolla, CA, USA  
Christine S. Liu & Carter R. Palmer
19. University of Connecticut, Storrs, CT, USA  
Marmar Moussa
20. Computer Science and Engineering Program, University of California, San Diego, La Jolla, CA, USA  
Sheng-Yong Niu
21. Department of Medical Biochemistry and Biophysics, Karolinska Institutet, Stockholm, Sweden  
Kimberly Siletti & Sten Linnarsson
22. McDonnell Genome Institute, Washington University School of Medicine, St Louis, MO, USA  
Xinxin Wang
23. Department of Physics, University of California, San Diego, La Jolla, CA, USA  
Fangming Xie
24. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, USA  
Alexander Dobin
25. Departments of Otorhinolaryngology, Anatomy and Neurobiology, University of Maryland School of Medicine, Baltimore, MD, USA  
Ronna Hertzano

26. Nash Family Department of Neuroscience and Friedman Brain Institute, Icahn School of Medicine at Mount Sinai, New York, NY, USA

Patrick R. Hof

27. Computer Graphics and Visualization Group, Delt University of Technology, Delft, The Netherlands

Thomas Höllt

28. Department of Physiology and Biophysics, Washington National Primate Research Center, University of Washington, Seattle, WA, USA

Gregory D. Horwitz

29. Department of Laboratory Medicine and Pathology, University of Washington, Seattle, WA, USA

C. Dirk Keene

30. Department of Neurological Surgery, University of Washington School of Medicine, Seattle, WA, USA

Andrew L. Ko

31. Regional Epilepsy Center, Harborview Medical Center, Seattle, WA, USA

Andrew L. Ko

32. Pattern Recognition and Bioinformatics group, Delft University of Technology, Delft, The Netherlands

Boudewijn P. Lelieveldt

33. Department of Human Genetics, University of California, Los Angeles, Los Angeles, CA, USA

Chongyuan Luo

34. Department of Cognitive Science, University of California, San Diego,  
La Jolla, CA, USA

Eran A. Mukamel

35. Broad Institute of MIT and Harvard, Cambridge, MA, USA

Aviv Regev, Evan Z. Macosko & Steven A. McCarroll

36. Department of Pathology, University of California, San Diego, CA,  
USA

Richard H. Scheuermann

37. Division of Vaccine Discovery, La Jolla Institute for Immunology, La  
Jolla, CA, USA

Richard H. Scheuermann

38. McGovern Institute for Brain Research, MIT, Cambridge, MA, USA

Guoping Feng

39. Department of Brain and Cognitive Sciences, MIT, Cambridge, MA,  
USA

Guoping Feng

40. Stanley Center for Psychiatric Research, Broad Institute of MIT and  
Harvard, Cambridge, MA, USA

Guoping Feng

41. Howard Hughes Medical Institute, The Salk Institute for Biological  
Studies, La Jolla, CA, USA

Joseph R. Ecker

## Corresponding authors

Correspondence to [Trygve E. Bakken](#) or [Ed S. Lein](#).

## 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

Bakken, T.E., Jorstad, N.L., Hu, Q. *et al.* Author Correction: Comparative cellular analysis of motor cortex in human, marmoset and mouse. *Nature* **604**, E8 (2022). <https://doi.org/10.1038/s41586-022-04562-y>

- Published: 22 March 2022
- Issue Date: 07 April 2022
- DOI: <https://doi.org/10.1038/s41586-022-04562-y>

### Share this article

Anyone you share the following link with will be able to read this content:

[Get shareable link](#)

Sorry, a shareable link is not currently available for this article.

[Copy to clipboard](#)

Provided by the Springer Nature SharedIt content-sharing initiative

---

This article was downloaded by **calibre** from <https://www.nature.com/articles/s41586-022-04562-y>.

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

[Download PDF](#)

- Publisher Correction
- [Published: 22 March 2022](#)

# Publisher Correction: Differential mechanisms underlie trace and delay conditioning in *Drosophila*

- [Dhruv Grover](#) ORCID: [orcid.org/0000-0001-5438-233X](#)<sup>1</sup>,
- [Jen-Yung Chen](#)<sup>1</sup>,
- [Jiayun Xie](#)<sup>1</sup>,
- [Jinfang Li](#) ORCID: [orcid.org/0000-0001-5754-8008](#)<sup>1</sup>,
- [Jean-Pierre Changeux](#)<sup>1,2,3</sup> &
- [Ralph J. Greenspan](#) ORCID: [orcid.org/0000-0002-6787-2845](#)<sup>1,4</sup>

[Nature](#) volume 604, page E9 (2022)

- 320 Accesses
- 2 Altmetric
- [Metrics details](#)

## Subjects

- [Attention](#)
- [Classical conditioning](#)
- [Cognitive control](#)
- [Neural circuits](#)
- [Working memory](#)

The [Original Article](#) was published on 16 February 2022

Correction to: *Nature* <https://doi.org/10.1038/s41586-022-04433-6>

Published online 16 February 2022

In the version of this article initially published, there were composition errors in units listed in the Methods. In the “Tethered-flight behaviour assay” section, first sentence of the second paragraph “2 cm” should have read “2 inch (or 5.08 cm),” while in the first sentence of the third paragraph of that section “2 cm × 2 cm” should have read “2 inch × 2 inch (or 5.08 cm × 5.08 cm).” In the “Two-photon calcium imaging” section, second sentence, “4 cm diameter, 3.5 cm height” should have read “4 inch diameter, 3.5 inch height (or 10.16 cm diameter, 8.89 cm height).” The changes have been made to the HTML and PDF versions of the article

## Author information

### Affiliations

1. Kavli Institute for Brain and Mind, University of California, San Diego, La Jolla, CA, USA

Dhruv Grover, Jen-Yung Chen, Jiayun Xie, Jinfang Li, Jean-Pierre Changeux & Ralph J. Greenspan

2. CNRS UMR 3571, Institut Pasteur, Paris, France

Jean-Pierre Changeux

3. College de France, Paris, France

Jean-Pierre Changeux

4. Division of Biological Sciences, University of California, San Diego, La Jolla, CA, USA

Ralph J. Greenspan

## Corresponding author

Correspondence to [Ralph J. Greenspan](#).

## Rights and permissions

[Reprints and Permissions](#)

## About this article

### Cite this article

Grover, D., Chen, JY., Xie, J. *et al.* Publisher Correction: Differential mechanisms underlie trace and delay conditioning in *Drosophila*. *Nature* **604**, E9 (2022). <https://doi.org/10.1038/s41586-022-04651-y>

- Published: 22 March 2022
- Issue Date: 07 April 2022
- DOI: <https://doi.org/10.1038/s41586-022-04651-y>

### Share this article

Anyone you share the following link with will be able to read this content:

[Get shareable link](#)

Sorry, a shareable link is not currently available for this article.

[Copy to clipboard](#)

Provided by the Springer Nature SharedIt content-sharing initiative

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

[Download PDF](#)

- Publisher Correction
- Open Access
- [Published: 22 March 2022](#)

# Publisher Correction: Structural basis for mismatch surveillance by CRISPR–Cas9

- [Jack P. K. Bravo](#)<sup>1,na1</sup>,
- [Mu-Sen Liu](#) [ORCID: orcid.org/0000-0002-3220-9330](#)<sup>1,na1</sup>,
- [Grace N. Hibshman](#)<sup>1,2</sup>,
- [Tyler L. Dangerfield](#)<sup>1,2</sup>,
- [Kyungseok Jung](#)<sup>1</sup>,
- [Ryan S. McCool](#) [ORCID: orcid.org/0000-0001-8833-8847](#)<sup>1,2</sup>,
- [Kenneth A. Johnson](#) [ORCID: orcid.org/0000-0002-6575-2823](#)<sup>1,2</sup> &
- [David W. Taylor](#) [ORCID: orcid.org/0000-0002-6198-1194](#)<sup>1,2,3,4</sup>

[Nature](#) volume **604**, page E10 (2022)

- 570 Accesses
- 2 Altmetric
- [Metrics details](#)

## Subjects

- [Cryoelectron microscopy](#)
- [DNA metabolism](#)
- [Enzyme mechanisms](#)

- [Genetic engineering](#)

The [Original Article](#) was published on 02 March 2022

Correction to: *Nature* <https://doi.org/10.1038/s41586-022-04470-1>  
Published online 2 March 2022

In the version of this article initially published, Extended Data Fig. 4 was an inadvertent duplicate of Extended Data Fig. 2. The image has been replaced in the HTML and PDF versions of the article

## Author information

### Author notes

1. These authors contributed equally: Jack P. K. Bravo, Mu-Sen Liu

## Affiliations

1. Department of Molecular Biosciences, University of Texas at Austin, Austin, TX, USA

Jack P. K. Bravo, Mu-Sen Liu, Grace N. Hibshman, Tyler L. Dangerfield, Kyungseok Jung, Ryan S. McCool, Kenneth A. Johnson & David W. Taylor

2. Interdisciplinary Life Sciences Graduate Programs, University of Texas at Austin, Austin, TX, USA

Grace N. Hibshman, Tyler L. Dangerfield, Ryan S. McCool, Kenneth A. Johnson & David W. Taylor

3. Center for Systems and Synthetic Biology, University of Texas at Austin, Austin, TX, USA

David W. Taylor

4. Livestrong Cancer Institutes, Dell Medical School, University of Texas at Austin, Austin, TX, USA

David W. Taylor

## Corresponding authors

Correspondence to [Kenneth A. Johnson](#) or [David W. Taylor](#).

## 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

Bravo, J.P.K., Liu, M.S., Hibshman, G.N. *et al.* Publisher Correction: Structural basis for mismatch surveillance by CRISPR–Cas9. *Nature* **604**, E10 (2022). <https://doi.org/10.1038/s41586-022-04655-8>

- Published: 22 March 2022

- Issue Date: 07 April 2022
- DOI: <https://doi.org/10.1038/s41586-022-04655-8>

## Share this article

Anyone you share the following link with will be able to read this content:

[Get shareable link](#)

Sorry, a shareable link is not currently available for this article.

[Copy to clipboard](#)

Provided by the Springer Nature SharedIt content-sharing initiative

---

This article was downloaded by **calibre** from <https://www.nature.com/articles/s41586-022-04655-8>

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

[Download PDF](#)

- Publisher Correction
- Open Access
- [Published: 23 March 2022](#)

# Publisher Correction: ER proteins decipher the tubulin code to regulate organelle distribution

- [Pengli Zheng](#) [ORCID: orcid.org/0000-0002-3797-4023<sup>1</sup>](#),
- [Christopher J. Obara<sup>2</sup>](#),
- [Ewa Szczesna<sup>3</sup>](#),
- [Jonathon Nixon-Abell](#) [ORCID: orcid.org/0000-0003-4169-0012<sup>1,2 nAff7</sup>](#),
- [Kishore K. Mahalingan<sup>3</sup>](#),
- [Antonina Roll-Mecak<sup>3,4</sup>](#),
- [Jennifer Lippincott-Schwartz](#) [ORCID: orcid.org/0000-0002-8601-3501<sup>2</sup>](#) &
- [Craig Blackstone](#) [ORCID: orcid.org/0000-0003-1261-9655<sup>1,5,6</sup>](#)

[Nature](#) volume 604, page E11 (2022)

- 464 Accesses
- 1 Altmetric
- [Metrics details](#)

## Subjects

- [Endoplasmic reticulum](#)

- [Microtubules](#)

The [Original Article](#) was published on 15 December 2021

Correction to: *Nature* <https://doi.org/10.1038/s41586-021-04204-9>

Published online 15 December 2021

In the Methods section of this Article, in the ‘CRISPR–Cas9 gene editing’ section, the sequences used as a homologous recombination template for the two proteins (CLIMP63 and calreticulin) should have been separated. The original Article has been corrected online

## Author information

### Author notes

1. Jonathon Nixon-Abell

Present address: Cambridge Institute for Medical Research,  
Cambridge, UK

### Affiliations

1. Cell Biology Section, Neurogenetics Branch, National Institute of Neurological Disorders and Stroke, National Institutes of Health, Bethesda, MD, USA

Pengli Zheng, Jonathon Nixon-Abell & Craig Blackstone

2. Janelia Research Campus, Howard Hughes Medical Institute, Ashburn, VA, USA

Christopher J. Obara, Jonathon Nixon-Abell & Jennifer Lippincott-Schwartz

3. Cell Biology and Biophysics Section, National Institute of Neurological Disorders and Stroke, National Institutes of Health, Bethesda, MD, USA

Ewa Szczesna, Kishore K. Mahalingan & Antonina Roll-Mecak

4. Biochemistry and Biophysics Center, National Heart, Lung and Blood Institute, National Institutes of Health, Bethesda, MD, USA

Antonina Roll-Mecak

5. MassGeneral Institute for Neurodegenerative Disease, Massachusetts General Hospital, Charlestown, MA, USA

Craig Blackstone

6. Department of Neurology, Massachusetts General Hospital and Harvard Medical School, Boston, MA, USA

Craig Blackstone

## Corresponding authors

Correspondence to [Pengli Zheng](#) or [Craig Blackstone](#).

## 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

Zheng, P., Obara, C.J., Szczesna, E. *et al.* Publisher Correction: ER proteins decipher the tubulin code to regulate organelle distribution. *Nature* **604**, E11 (2022). <https://doi.org/10.1038/s41586-022-04656-7>

- Published: 23 March 2022
- Issue Date: 07 April 2022
- DOI: <https://doi.org/10.1038/s41586-022-04656-7>

### Share this article

Anyone you share the following link with will be able to read this content:

[Get shareable link](#)

Sorry, a shareable link is not currently available for this article.

[Copy to clipboard](#)

Provided by the Springer Nature SharedIt content-sharing initiative

---

This article was downloaded by **calibre** from <https://www.nature.com/articles/s41586-022-04656-7>

[Download PDF](#)

- Retraction Note
- [Published: 24 February 2022](#)

# Retraction Note: A 10 per cent increase in global land evapotranspiration from 2003 to 2019

- [Madeleine Pascolini-Campbell](#) ORCID: [orcid.org/0000-0002-6449-0841](https://orcid.org/0000-0002-6449-0841)<sup>1</sup>,
- [John T. Reager](#) ORCID: [orcid.org/0000-0001-7575-2520](https://orcid.org/0000-0001-7575-2520)<sup>1</sup>,
- [Hrishikesh A. Chandanpurkar](#) ORCID: [orcid.org/0000-0002-7573-8056](https://orcid.org/0000-0002-7573-8056)<sup>1</sup> &
- [Matthew Rodell](#) ORCID: [orcid.org/0000-0003-0106-7437](https://orcid.org/0000-0003-0106-7437)<sup>2</sup>

[Nature](#) volume 604, page 202 (2022)

- 3749 Accesses
- 18 Altmetric
- [Metrics details](#)

## Subjects

- [Climate change](#)
- [Climate sciences](#)
- [Hydrology](#)

The [Original Article](#) was published on 26 May 2021

Correction to: *Nature* <https://doi.org/10.1038/s41586-021-03503-5>

Published online 26 May 2021

In this article, we calculated global land evapotranspiration for 2003 to 2019 using a mass-balance approach. To do this, we calculated evapotranspiration as the residual of the water balance, using an ensemble of datasets for precipitation, discharge and total water storage change. We made an error in calculating the global mean precipitation: we used arithmetic averaging to calculate the mean, instead of calculating a spatially weighted mean to account for the changing grid box size with latitude. As a result, the magnitudes of the global mean precipitation time series were underestimated. This impacted the subsequent calculation of global mean evapotranspiration, resulting in the mean evapotranspiration values being underestimated and altering some results. We are therefore retracting this article. We thank Ning Ma and others for bringing this error to our attention.

## Author information

### Affiliations

1. NASA Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA, USA

Madeleine Pascolini-Campbell, John T. Reager & Hrishikesh A. Chandanpurkar

2. NASA Goddard Space Flight Center, Greenbelt, MD, USA

Matthew Rodell

### Corresponding author

Correspondence to [Madeleine Pascolini-Campbell](#).

# Rights and permissions

[Reprints and Permissions](#)

## About this article

### Cite this article

Pascolini-Campbell, M., Reager, J.T., Chandanpurkar, H.A. *et al.* Retraction Note: A 10 per cent increase in global land evapotranspiration from 2003 to 2019. *Nature* **604**, 202 (2022). <https://doi.org/10.1038/s41586-022-04525-3>

- Published: 24 February 2022
- Issue Date: 07 April 2022
- DOI: <https://doi.org/10.1038/s41586-022-04525-3>

### Share this article

Anyone you share the following link with will be able to read this content:

[Get shareable link](#)

Sorry, a shareable link is not currently available for this article.

[Copy to clipboard](#)

Provided by the Springer Nature SharedIt content-sharing initiative

---

This article was downloaded by **calibre** from <https://www.nature.com/articles/s41586-022-04525-3>

# Career Guide