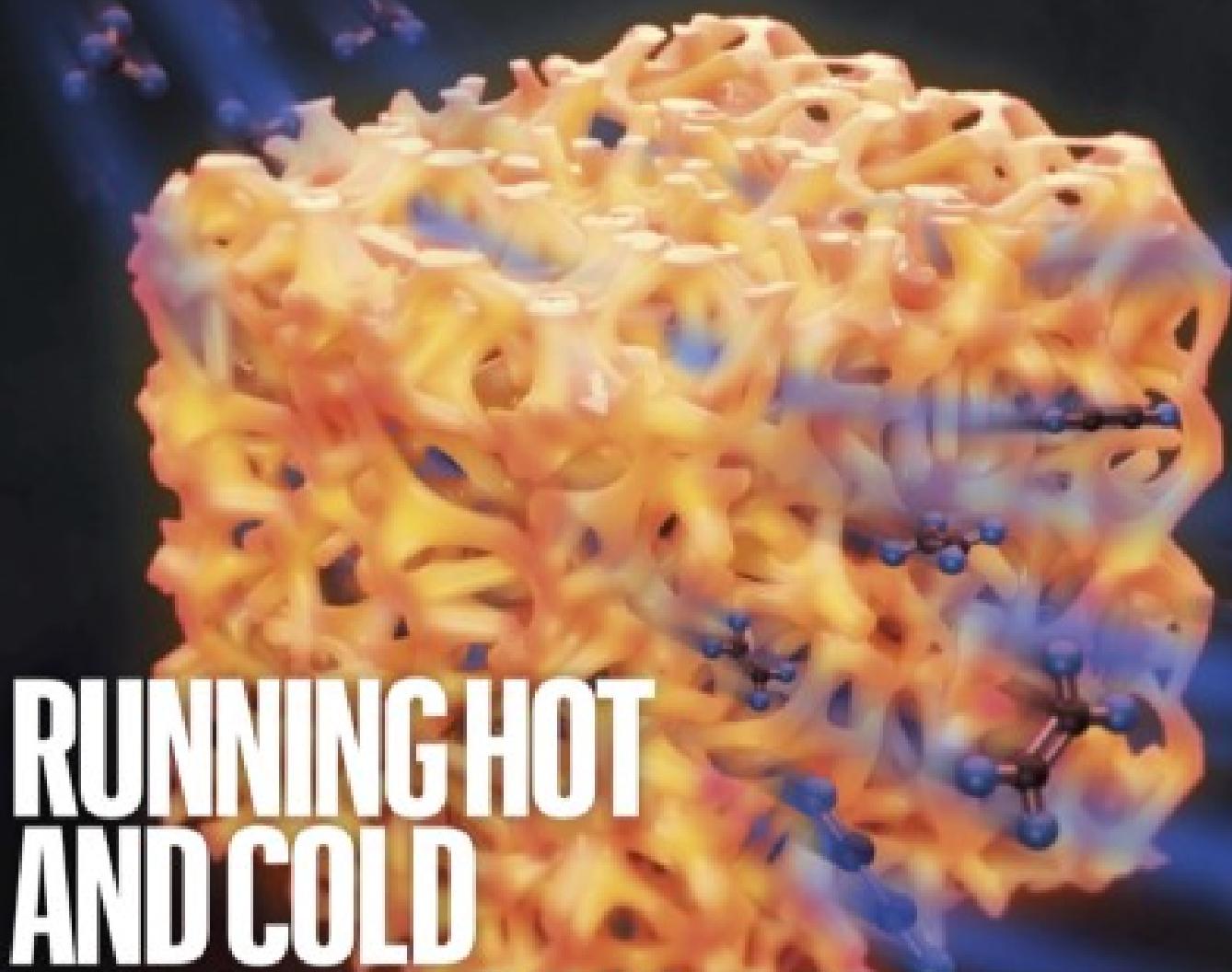


nature



RUNNING HOT AND COLD

Rapid heating and quenching offers performance improvements for thermochemical syntheses

Russian invasion

How some of Ukraine's scientists are trying to cope amid the war

Pandemic prevention

Global agreements must defend against pathogen spillover

Aided recall

Old mice get memory boost from youthful cerebrospinal fluid

Nature.2022.05.21

[Sat, 21 May 2022]

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

This Week

- **The toll of menopause: how universities can help** [11 May 2022]
Editorial • Some women are leaving science because employers are failing to support them during this stage of life. That can't be right.
- **Nature journals raise the bar on sex and gender reporting in research** [18 May 2022]
Editorial • Authors will be prompted to provide details on how sex and gender were considered in study design.
- **How climate law can help to prevent the next pandemic** [17 May 2022]
World View • Countries seeking to effect real changes in global health can learn from climate treaties.

- EDITORIAL
- 11 May 2022

The toll of menopause: how universities can help

Some women are leaving science because employers are failing to support them during this stage of life. That can't be right.



Menopause coincides with a time when people move into more senior roles. Credit: Luis Alvarez/Getty

Half of all the people on Earth will go through menopause. It is a natural part of ageing, affecting the majority of women, as well as some trans men

and non-binary people. Levels of hormones including oestrogen, progesterone and testosterone decline, causing symptoms that can last a decade or more. Unfortunately, it is only in the past decade or so that its effects on women's lives and careers have been the focus of more than a handful of studies.

Women don't need telling that menopause symptoms, such as insomnia, fatigue and difficulty focusing, can have a major impact on their lives. Researchers are now discovering that such symptoms also disrupt a significant number of careers. In Japan last year, [a study of thousands of people](#) — the country's first to focus on menopause and work — found that 'menopausal loss' affected one-fifth of women experiencing menopause, who quit, turned down promotions, reduced their working hours or were demoted as a result of their symptoms.

Menopause often comes at a time when people move into more senior, more demanding roles. In research, as in other careers, this coincidence is almost certainly causing some to reconsider their career ambitions, adding yet another drain to the 'leaky pipeline' of women in science. Women going through this stage of life shouldn't have to row back or leave careers because of a lack of support from employers, as [we report in a Careers Feature](#). Everyone must make their own choices, but no one should feel forced into any decision because of an unwelcoming workplace.

There are many ways in which research workplaces can support staff going through menopause, as the authors of an essay last month in *BDJ Team* suggest ([J. A. Bell et al. BDJ Team 9, 24–26; 2022](#)). Offering flexible working hours and accepting menopause symptoms as a valid reason to take sick leave are good places to start. Some workplaces are also creating quiet spaces for staff; this not only helps those experiencing noise sensitivity, hot flushes and other symptoms of menopause, but also benefits those seeking quiet time for other reasons.

In some countries, people have established peer networks — either in-person or virtual — to support women going through menopause. One example is [Red Hot Mamas](#) in Canada and the United States. Another is [menopause 'cafés'](#), which began in the United Kingdom and offer informal gatherings

that anyone, regardless of sex or gender, can attend to learn about and discuss menopause over a bite to eat and a drink.

The Royal Society in London, the world's oldest science academy, launched a menopause support group for staff last year and is drawing up guidelines to ensure that women know how to access support, such as adjustments to roles, working hours or technology. The University of Nottingham, which has campuses in the United Kingdom, China and Malaysia, also has guidelines to help managers support staff going through menopause. This could include offering flexible working hours or better ventilation — by providing desk fans, for example, or, better still, using innovative approaches to interior architecture and design.

Workplaces can also offer advice on how to manage symptoms or provide information on where to obtain such advice — although more research is needed in this area, too.

However, such support needs to be offered with care. If institutional policies and support programmes are drawn up without involving the individuals most affected, they risk making those they aim to support feel self-conscious and stigmatized. Several women contacted by *Nature* expressed concern that highlighting the challenges some women face during menopause could put academic employers off hiring older women. Those fears are understandable, and there should be some ground rules for the discussion. Institutions should never force anyone to disclose their menopause status, nor require them to engage in discussions of the topic if they prefer not to.

Doing nothing should never be an option. Many employers (including Springer Nature) are pledging workplace support for people experiencing menopause. These pledges include ensuring that they are listened to, should they approach their managers, and that practical support is available. But these are just first steps. There is more that all employers, including those in the scientific and research space, can do.

The research community also needs to devote more attention and resources to studying the impact of menopause on careers everywhere, not just in high-income countries. And those organizations that have not yet started to address the difficulties that menopause can pose for working life need to do

so now. It's time for the stigma around menopause to be lifted. Doing so will make research a better place to work for everyone.

Nature **605**, 395-396 (2022)

doi: <https://doi.org/10.1038/d41586-022-01297-8>

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-022-01297-8>

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

- EDITORIAL
- 18 May 2022

Nature journals raise the bar on sex and gender reporting in research

Authors will be prompted to provide details on how sex and gender were considered in study design.



Many research studies don't account for sex and gender.Credit: Getty

In late 2020, the European Commission announced that its research-grant recipients would need to incorporate analyses of sex and gender in their study design. This could include disaggregating data by sex when examining

cells, or considering how a technology might perpetuate gender stereotypes. Back then, *Nature* wrote that this was a significant step and urged other funders to follow suit (see [*Nature* 588, 196; 2020](#)). At the same time, we said that publishers, too, have a role in encouraging sex and gender reporting. The responsibility does not lie only with funders.

Some journals have encouraged reporting of sex and gender analyses for years, and the number of research studies that include such data has increased substantially in the past decade. But gaps remain — especially insufficient reporting of data disaggregated by sex and gender^{1–3}.

To remedy this, from now on, researchers who submit papers to a subset of *Nature* Portfolio journals (see list at go.nature.com/3mcu0zj) will be prompted to state whether and how sex and gender were considered in their study design, or to indicate that no sex and gender analyses were carried out, and clarify why. They should note in the title and/or abstract if findings apply to only one sex or gender.

They will also be asked to provide data disaggregated by sex and gender where this information has been collected, and informed consent for reporting and sharing individual-level data has been obtained. The changes apply to studies with human participants, on other vertebrates or on cell lines, in which sex and gender is an appropriate consideration.

At the same time, we're urging care and caution in communicating findings about sex and gender, to avoid research findings having inadvertent and harmful effects, especially where there is the potential for societal and public-policy impact. More details about these changes can be found at go.nature.com/3mcu0zj. They are part of the SAGER (Sex and Gender Equity in Research) guidelines⁴.

In addition, from 1 June, four journals — *Nature Cancer*, *Nature Communications*, *Nature Medicine* and *Nature Metabolism* — will be raising awareness of the updated recommendations in letters to authors and reviewers during peer review. The aim here is to improve understanding of the degree to which sex and gender reporting is already part of study design, data collection and analysis in the research these journals publish. The

journals will also evaluate author and reviewer reception of the changes so that we can iterate on them as we learn through experience.

The new measures are needed because research is still mostly failing to account for sex and gender in study design, sometimes with catastrophic results. Between 1997 and 2001, ten prescription drugs were withdrawn from use in the United States; eight of these were reported to have worse side effects in women than in men (we recognize that not everyone fits into these categories). These differences had probably been missed, in part, because of insufficient or inappropriate analysis of data on sex differences during clinical trials.

By introducing these changes, we aim to promote transparency in study design and, ultimately, make findings more accurate. Over time, we hope to see integration of sex and gender analysis in study design by default.

Nature **605**, 396 (2022)

doi: <https://doi.org/10.1038/d41586-022-01218-9>

References

1. Woitowich, N. C., Beery, A. & Woodruff, T. *eLife* **9**, e56344 (2020).
2. Rechlin, R. K., Splinter, T. F. L., Hodges, T. E., Albert, A. Y. & Galea, L. A. M. *Nature Commun.* **13**, 2137 (2021).
3. Brady, E., Wullum Nielsen, M., Andersen, J. P. & Oertelt-Prigione, S. *Nature Commun.* **12**, 4015 (2021).
4. Heidari, S. *et al. Res. Integr. Peer Rev.* **1**, 2 (2016).

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-022-01218-9>

- WORLD VIEW
- 17 May 2022

How climate law can help to prevent the next pandemic



Countries seeking to effect real changes in global health can learn from climate treaties.

- [Alexandra L. Phelan](#) 0

Two of the biggest crises humanity faces — pandemics and climate change — are intertwined. Climate change boosts many health risks, including the likelihood that new viruses will spill over to cause dangerous outbreaks. But although efforts to control climate change are bolstered by a network of international treaties and legal agreements, such tools have not been fully applied to global health. Those gathering for the World Health Assembly in Geneva later this month should push to change this.

I am a lawyer and researcher specializing in the governance of pandemics and climate change. I've seen international lawmaking produce political commitments such as national emissions targets. Since 1992, a network of treaties, including the United Nations Framework Convention on Climate Change (UNFCCC), have erected protocols to build consensus. Progress has been slow, but real.

This is exemplified by the 2015 Paris Agreement. Yes, it is inadequate: countries set their own, non-binding targets. Nonetheless, presidents and prime ministers have made pledges to reduce emissions and are now publicly accountable. Countries have also explicitly recognized that changes in climate (such as extreme heat, droughts and flooding) can infringe on the right to health by contributing to crop failure, infectious diseases and other disasters. The agreement captured political momentum to get countries to make prominent commitments that increase over time.

Pandemics lack any equivalent scaffolding to support complex global cooperation. The International Health Regulations, which were last reworked in 2005, have significant gaps. Although they are legally binding, enforcement is weak and they are largely ignored. The global health community, often loath to seem 'political', has underused the potential of international law to build compliance norms.

In 2021, the World Health Organization's member states established a formal negotiating body to explore international law for pandemic prevention, preparedness and response. When it sought input last month, I made two points. First, that pandemic law should engage countries' existing legal obligations by acknowledging how climate change will exacerbate outbreaks. Second, that a pandemic treaty could be modelled on climate law to make countries transparent in and accountable for reaching commitments. The specifics — of viral surveillance, information sharing and so on — are less urgent than the process, with one exception. Pandemic law should learn from failures of climate law, and ensure that attention is paid to justice and equity across and within countries.

The UNFCCC was written to spur action. A framework convention is a treaty that sets out high-level, legally binding principles and obligations to support faster negotiation and adoption. It can distil political momentum

into national commitments, including governance structures and processes. It also allows for protocols, such as the Kyoto Protocol or Paris Agreement, that can be refined in parallel or over time, so negotiators can build on past progress and create detailed obligations for specific issues, such as technology transfer or equitable vaccine distribution.

The power of the UNFCCC lies in how it establishes institutions and processes to support collective action and accountability. Conferences of parties (COPs) are the clearest example. Recall how COP26 in Glasgow last year captured the world's attention and pushed leaders to set more-ambitious goals. COPs exist to assess, clarify and reiterate obligations. Non-governmental organizations, advocacy organizations and other components of civil society use COPs to hold governments accountable. Citizens can also hold governments to account for failures to take sufficient action, as is happening with class action climate lawsuits in more than 35 nations. Indeed, the fact that there are so many strong global public-health organizations could make such mechanisms more powerful.

The key point is not to look at any single treaty, accord or policy as the outcome — there have been so many disappointments. We should look at how these mechanisms function together. Financing and capacity building are key to efficacy under the UNFCCC, the Paris Agreement and decisions of the COPs. Accountability and transparency is also crucial. Pandemic law must, for example, protect the rights of whistle-blowers, including health workers.

Better than punitive measures, which can erode cooperation, are mechanisms to encourage compliance. For instance, the compliance committee established by the Paris Agreement helps countries to make progress on emissions targets by identifying non-compliance, providing expert guidance on requirements and deadlines, and enforcing plans.

The Intergovernmental Panel on Climate Change produces regular reports that provide reliable updates and syntheses of available evidence. Similarly, an effective pandemic treaty would establish an independent process for collecting and synthesizing scientific evidence for preparedness and response. This would guide investments in building capacity and developing technologies, and would inform policy for mitigating outbreaks.

International climate law is far from sufficient: countries have not yet lowered emissions enough to avoid a hotter and sicker world. But they have enabled climate action. Any pandemic treaty will be imperfect. But taking difficult, imperfect steps is the only way to build momentum.

Nature **605**, 397 (2022)

doi: <https://doi.org/10.1038/d41586-022-01339-1>

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-022-01339-1>

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

News in Focus

- **[Parasite names, mouse rejuvenation and toxic sunscreen](#)** [18 May 2022]
News Round-Up • The latest science news, in brief.
- **[Black hole at the centre of our Galaxy imaged for the first time](#)** [12 May 2022]
News • The Event Horizon Telescope network has captured the second-ever direct image of a black hole — called Sagittarius A* — at the centre of the Milky Way.
- **[China expands control over genetic data used in scientific research](#)** [06 May 2022]
News • Guidelines released this year are the latest regulations to protect China's genetic resources, but some scientists say they are making collaborations harder.
- **[This Arctic town wants to make renewable energy work at the top of the world](#)** [06 May 2022]
News • Partnering with a northern settlement in Greenland, researchers are designing wind and solar devices that can survive and thrive in extreme conditions.
- **[Coronavirus ‘ghosts’ found lingering in the gut](#)** [11 May 2022]
News • Scientists are studying whether long COVID could be linked to viral fragments found in the body months after initial infection.
- **[Origin of life theory involving RNA–protein hybrid gets new support](#)** [11 May 2022]
News • Structure that links amino acids suggests that early organisms could have been based on an RNA–protein mix.
- **[The pandemic’s true health cost: how much of our lives has COVID stolen?](#)** [18 May 2022]
News Feature • Researchers are trying to calculate how many years have been lost to disability and death.
- **[How three Ukrainian scientists are surviving Russia’s brutal war](#)** [06 May 2022]
News Feature • Helping, healing and fighting: researchers have become refugees, soldiers and activists in the face of a horrifying conflict.

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

- NEWS ROUND-UP
- 18 May 2022

Parasite names, mouse rejuvenation and toxic sunscreen

The latest science news, in brief.



Young cerebrospinal fluid probably improves the conductivity of the neurons in ageing mice.Credit: Qilai Shen/Bloomberg/Getty

Young brain fluid improves memory in old mice

Cerebrospinal fluid (CSF) from young mice can improve memory function in older mice, researchers report in *Nature* ([T. Iram et al. *Nature* 605, 509–515; 2022](#)).

A direct brain infusion of young CSF [probably improves the conductivity of the neurons in ageing mice](#), which improves the process of making and recalling memories.

CSF is a cocktail of essential ions and nutrients that cushions the brain and spinal cord and is essential for normal brain development. But as mammals age, CSF loses some of its punch. Those changes might affect cells related to memory, says co-author Tal Iram, a neuroscientist at Stanford University in California.

The researchers found that young CSF helps ageing mice to generate more early-stage oligodendrocytes, cells in the brain that produce the insulating sheath around nerve projections and help to maintain brain function.

The team suggest that the improvements are largely due to a specific protein in the fluid.

“This is super exciting from the perspective of basic science, but also looking towards therapeutic applications,” says Maria Lehtinen, a neurobiologist at Boston Children’s Hospital in Massachusetts.

Gender bias worms its way into parasite naming

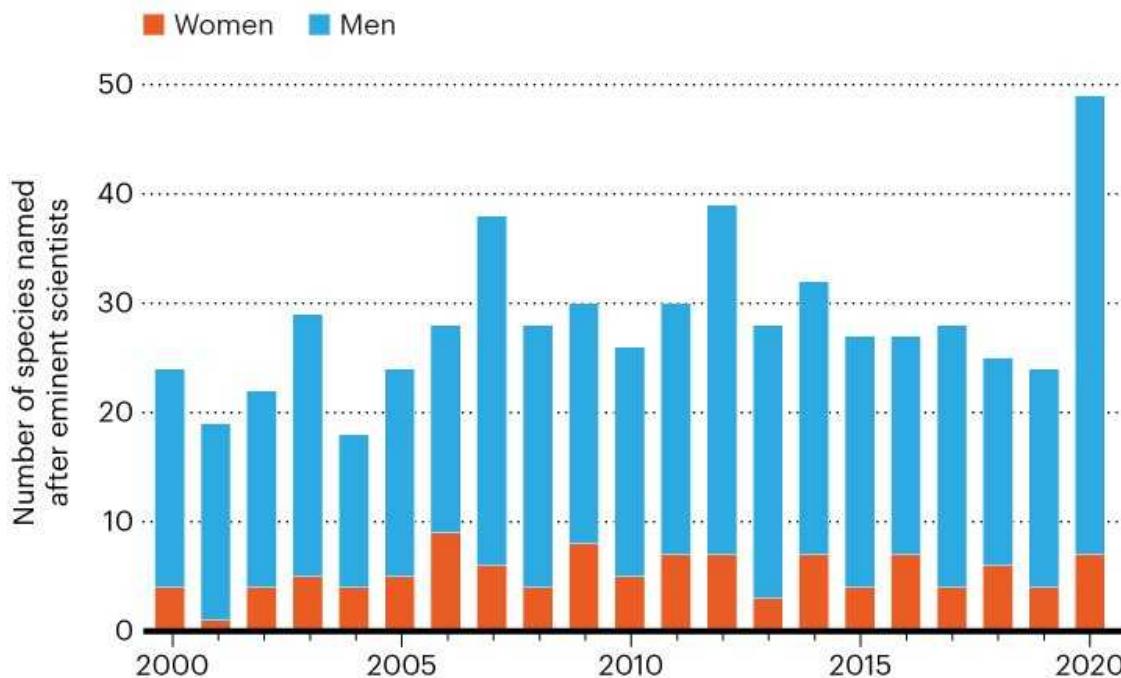
A study examining the names of nearly 3,000 species of parasitic worm discovered in the past 20 years reveals a markedly higher proportion named after male scientists than after female scientists — and a [growing appetite for immortalizing friends and family members in scientific names](#).

Robert Poulin, an ecological parasitologist at the University of Otago in Dunedin, New Zealand, and his colleagues combed through papers published between 2000 and 2020 that describe roughly 2,900 new species of parasitic worm ([R. Poulin et al. *Proc. R. Soc. B* <https://doi.org/htqn>; 2022](#)). The team found that well over 1,500 species were named after their

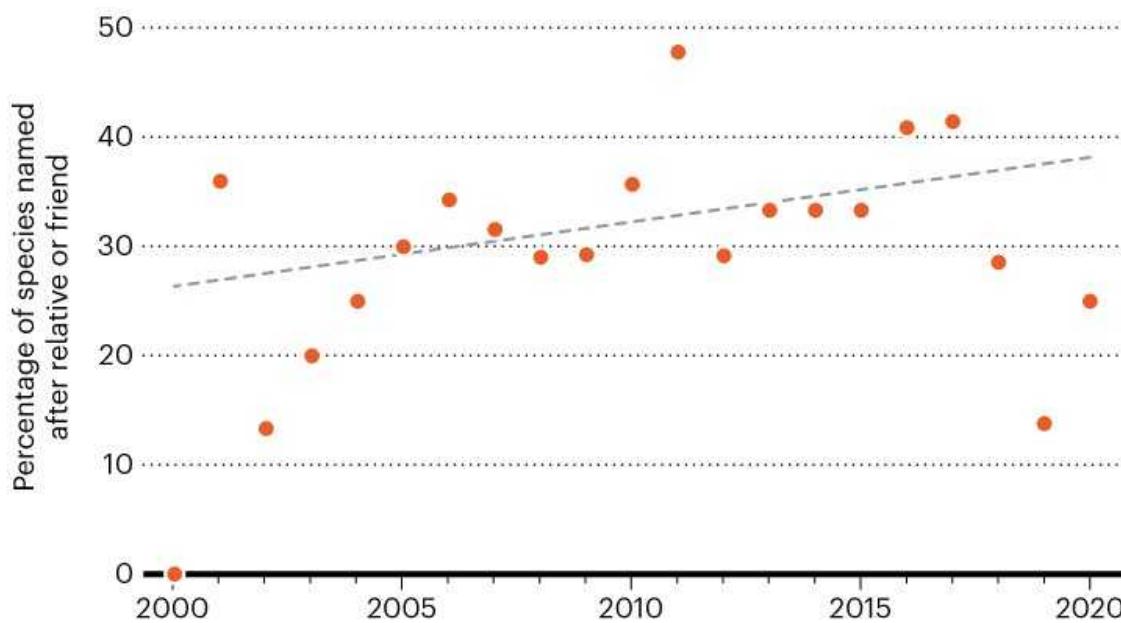
host organism, where they were found or a prominent feature of their anatomy.

PARASITE NAME GAME

From 2000 to 2020, nearly 600 newly described parasites were named after an eminent scientist — only 19% of them women. The proportion named after women has remained essentially static over those two decades.



Over the same time period, the fraction of parasite species named after an author's personal connections — friends, families and even pets — rose.



Source: R. Poulin *et al.* *Proc. R. Soc. B* <https://doi.org/htqn> (2022)

Many others were named after people, ranging from technical assistants to prominent politicians. But just 19% of the 596 species named after eminent scientists were named after women, a percentage that barely changed over the decades (see ‘Parasite name game’). Poulin and his colleagues also noticed an upward trend in the number of parasites named after friends, family members and even pets of the scientists who formally described them. This practice should be discouraged, Poulin argues.



Sea anemones turn oxybenzone into a light-activated agent that can bleach and kill corals. Credit: Georgette Douwma/Getty

Anemones suggest why sunscreen turns toxic in sea

A common but controversial sunscreen ingredient that is thought to harm corals might do so because of a [chemical reaction that causes it to damage cells in the presence of ultraviolet light](#).

Researchers have discovered that sea anemones, which are similar to corals, make the sun-blocking molecule oxybenzone water-soluble by tacking a sugar onto it. This inadvertently turns oxybenzone into a molecule that — instead of blocking UV light — is activated by sunlight to produce free radicals that can bleach and kill corals. The animals “convert a sunscreen into something that’s essentially the opposite of a sunscreen”, says Djordje Vuckovic, an environmental engineer at Stanford University in California.

It’s not clear how closely these laboratory-based studies mimic the reality of reef ecosystems. The concentration of oxybenzone at a coral reef can vary widely, depending on factors such as tourist activity and water conditions. And other factors threaten the health of coral reefs; these include climate change, ocean acidification, coastal pollution and overfishing. The study, published on 5 May ([D. Vuckovic *et al. Science* 376, 644–648; 2022](#)) does not show where oxybenzone ranks in the list.

Nature **605**, 401 (2022)

doi: <https://doi.org/10.1038/d41586-022-01340-8>

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-022-01340-8>

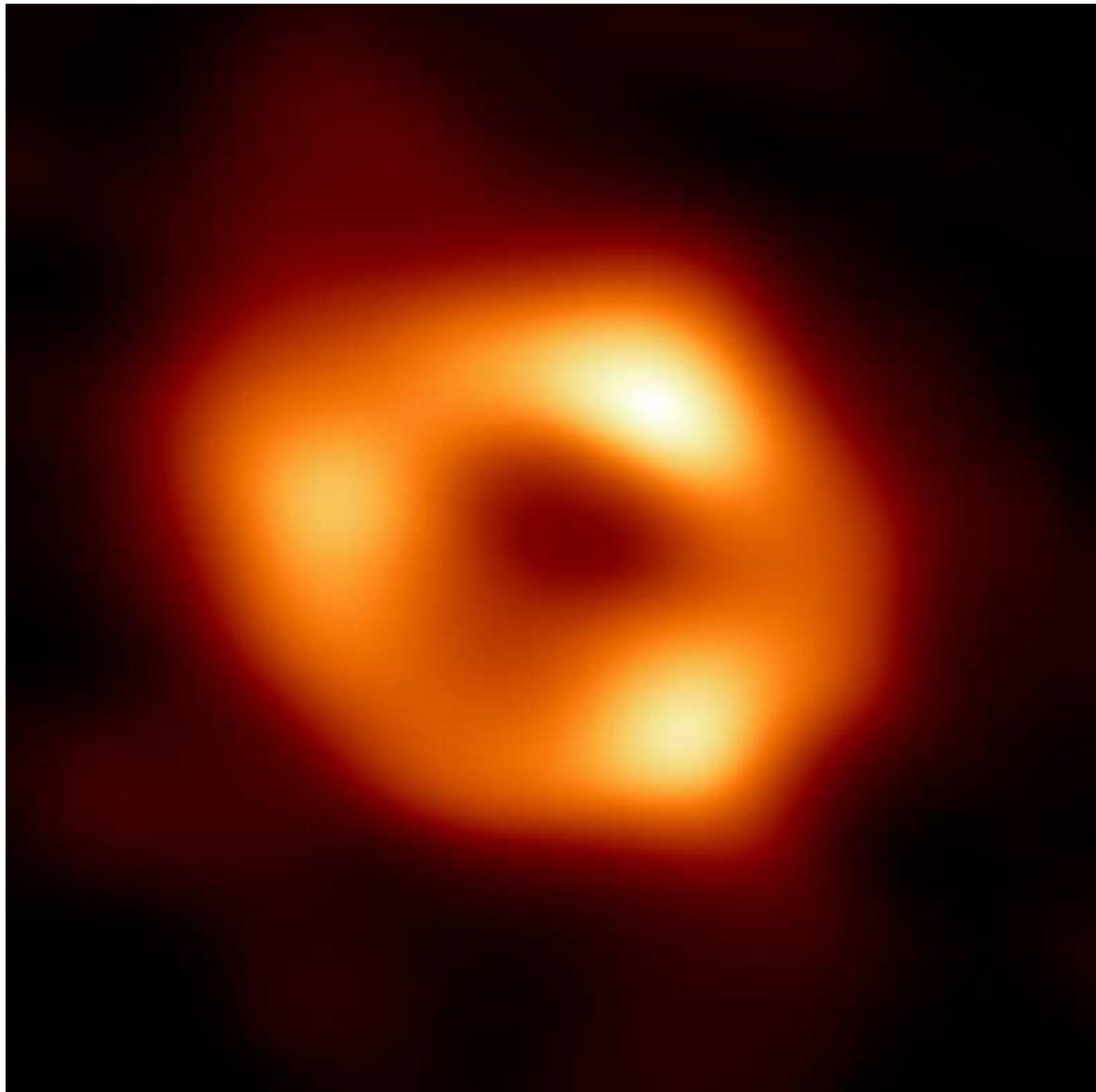
| [Section menu](#) | [Main menu](#) |

- NEWS
- 12 May 2022
- Correction [20 May 2022](#)

Black hole at the centre of our Galaxy imaged for the first time

The Event Horizon Telescope network has captured the second-ever direct image of a black hole — called Sagittarius A* — at the centre of the Milky Way.

- [Davide Castelvecchi](#)



The second-ever direct image of a black hole — Sagittarius A*, at the centre of the Milky Way.Credit: Event Horizon Telescope collaboration

Radioastronomers have imaged the supermassive black hole at the centre of the Milky Way. It is only the second-ever direct image of a black hole, after the same team unveiled a [historic picture of a more distant black hole](#) in 2019.

The long-awaited results, presented today by the Event Horizon Telescope (EHT) collaboration, show an image reminiscent of the earlier one: a ring of radiation surrounds a darker disk of precisely the size that was predicted

from indirect observations and from Albert Einstein's general theory of relativity.

"Today, right this moment, we have direct evidence that this object is a black hole," said Sara Issaoun, an astrophysicist at the Harvard and Smithsonian Center for Astrophysics in Cambridge, Massachusetts, at a press conference in Garching, Germany. The team published its results in [a special issue of The Astrophysical Journal Letters](#)¹.

"We've been working on this for so long, every once in a while you have to pinch yourself and remember that this is the black hole at the centre of our Universe," said EHT team member Katie Bouman, a computational-imaging researcher at the California Institute of Technology in Pasadena, at a press conference in Washington DC. "I mean, what's more cool than seeing the black hole at the centre of the Milky Way?"

Black-hole observations

During five nights in April 2017, the EHT collaboration used eight observatories across the world to collect data from both the Milky Way's black hole — called Sagittarius A*, after the constellation in which it is found — and M87*, the one at the centre of the galaxy M87.

The observatory locations ranged from Spain to the South Pole and from Chile to Hawaii. They collected nearly 4 petabytes (4,000 terabytes) of data, which was too much to be sent over the Internet and had to be carried by aeroplane on hard disks.

The EHT researchers [unveiled their image of M87*](#) in 2019, showing the first direct evidence of an event horizon, the spherical surface that shrouds a black hole's interior.

But the Sagittarius A* data were more challenging to analyse. The two black holes have roughly the same apparent size in the sky, because M87* is nearly 2,000 times farther away but about 1,600 times larger. Any [blobs of matter spiralling around M87*](#) are covering much larger distances — larger than Pluto's orbit around the Sun — and the radiation they emit is essentially

constant over short time scales. But Sagittarius A* can change quickly, even over the few hours that the EHT observes it every day. “In M87*, we saw very little variation within a week,” says Heino Falcke, an astrophysicist at Radboud University in Nijmegen, the Netherlands, and a co-founder of the EHT collaboration. “Sagittarius A* varies on time scales of 5 to 15 minutes.”

Because of this variability, the EHT team generated not one image of Sagittarius A*, but thousands — and the image unveiled today is the result of a lot of processing. “By averaging them together, we are able to emphasize common features,” says EHT member José Gómez, at the Institute of Astrophysics of Andalusia in Granada, Spain.

In addition to showing a ring of radiation around a darker disk, the resulting image contained three brighter ‘knots’. “We see knots in all the images we created,” says Issaoun, but each had the knots in different places. The averaged knots that appear in the image are probably artefacts of the interferometry technique used by the EHT, she adds. It reconstructs images from an idealized Earth-sized radio dish — but one in which only tiny shards of the dish are able to take data at any given time.

The appearance is different from that of M87*, for which the brighter region in the image had more of a half-moon shape, which could indicate a denser blob of matter being [accelerated along the direction of the line of sight](#).

The project’s next aim is to generate a movie of the black hole to learn more about its physical properties, Feryal Özel, an astrophysicist at the University of Arizona in Tucson, told reporters at the Washington DC press conference.

The EHT team conducted supercomputer simulations to compare with their data and concluded that Sagittarius A* is probably rotating anticlockwise along an axis that roughly points along the line of sight to Earth, said Gómez.

“What blows my mind is that we’re seeing it face-on,” says Regina Caputo, an astrophysicist at NASA’s Goddard Space Flight Center in Greenbelt, Maryland. NASA’s Fermi Gamma-ray Space Telescope, which Caputo works with, had previously detected giant glowing features above and below

the centre of the galaxy, which could have been produced by Sagittarius A* during periods of intense activity in the past. But those features, known as [Fermi bubbles](#), seem to require matter to swirl around the black hole edge-on, rather than face-on, as seen from Earth.

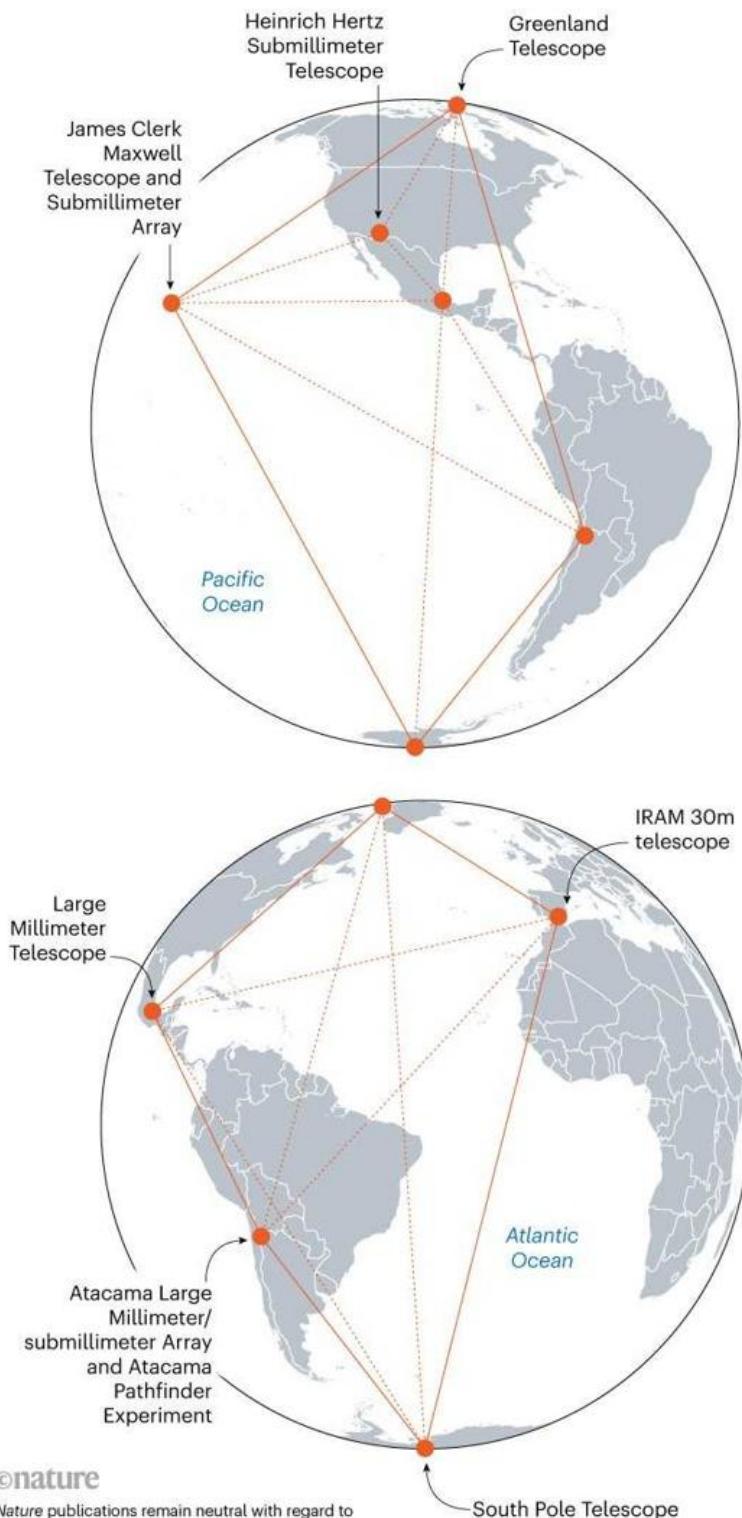
Extremely massive object

The first hints of the existence of Sagittarius A* were seen in the 1970s, when radioastronomers discovered a seemingly point-like radio source in the central region of the Galaxy.

The source turned out to be unusually dim, dimmer than an average star. Still, decades-long observations of the motions of nearby stars revealed that the object was extremely massive: using the most recent observations, scientists have calculated that it is 4.15 million times the mass of the Sun, give or take 0.3%. These calculations, done by tracking how stars orbit Sagittarius A*, provided strong evidence that the radio source is so massive and dense that it could be nothing but a black hole. That work earned Andrea Ghez, an astrophysicist at the University of California, Los Angeles, and Reinhard Genzel, director of the Max Planck Institute for Extraterrestrial Physics in Garching, Germany, a share of the [2020 Nobel Prize in Physics](#). (The size of the dark shadow in the EHT image suggests that the black hole weighs around 4 million solar masses, which is strikingly consistent with the earlier estimates, although not as precise.)

GLOBAL EFFORT

The Event Horizon Telescope combined signals from eight radio observatories across the world. Together, the observatories have a resolving power equivalent to a telescope almost the size of Earth.



©nature

Nature publications remain neutral with regard to contested jurisdictional claims in published maps.

Source: Event Horizon Telescope

Sagittarius A* is practically invisible to optical telescopes because of the dust and gas on the galactic disk. But beginning in the late 1990s, Falcke and others realized that the shadow of the black hole might be large enough to be imaged with short radio waves, which can pierce that veil. But researchers calculated that doing so would require a telescope the size of Earth. Fortunately, a technique called interferometry could help. It involves simultaneously pointing multiple far-away telescopes at the same object. Effectively, the telescopes work like shards of one big dish (see ‘Global effort’).

The first attempts to observe Sagittarius A* with interferometry used relatively long 7-millimetre radio waves and observatories a few thousand kilometres apart. All astronomers could see was a blurred spot.

Teams across the world then refined their techniques and retrofitted major observatories that were added to the network. In particular, researchers adapted the South Pole Telescope and the US\$1.4-billion Atacama Large Millimeter/submillimeter Array in Chile to do the work.

Then in 2015, groups joined forces as the EHT collaboration. Their 2017 observation campaign was the first to span distances long enough to resolve details such as the size of Sagittarius A*.

Future plans

The EHT collected more data in 2018 but [cancelled their planned observation campaigns](#) in 2019 and 2020. They resumed observations in 2021 and 2022, with an improved network and more sophisticated instruments.

Remo Tilanus, an EHT member at the University of Arizona in Tucson, says the team’s latest observations, in March, recorded signals at twice the 2017 rate, most of them at 0.87-millimetre wavelength — which should help to increase the resolution of the resulting images.

Researchers hope to find out whether Sagittarius A* has jets. Many black holes, including M87*, display two beams of matter rapidly shooting out in opposite directions, thought to be a result of the intense heating of infalling gas [and powered by the black hole's spin](#). Sagittarius A* might have had large jets in the past — as heated clouds of matter above and below the galactic centre suggest. Its jets would now be much weaker, but their presence could reveal important details about our Galaxy's history.

“These jets can inhibit or induce star formation, they can move the chemical elements around,” and affect the evolution of an entire galaxy, says Falcke. “And we’re now looking at where it’s happening.”

Nature **605**, 403-404 (2022)

doi: <https://doi.org/10.1038/d41586-022-01320-y>

Additional reporting by Freda Kreier.

Updates & Corrections

- **Correction 20 May 2022:** An earlier version of this story incorrectly described Katie Bouman as a former EHT team member. She is a current member of the EHT team.

References

1. Akiyama, K. *et al. Astrophys. J. Lett.* **930**, L12 (2022).

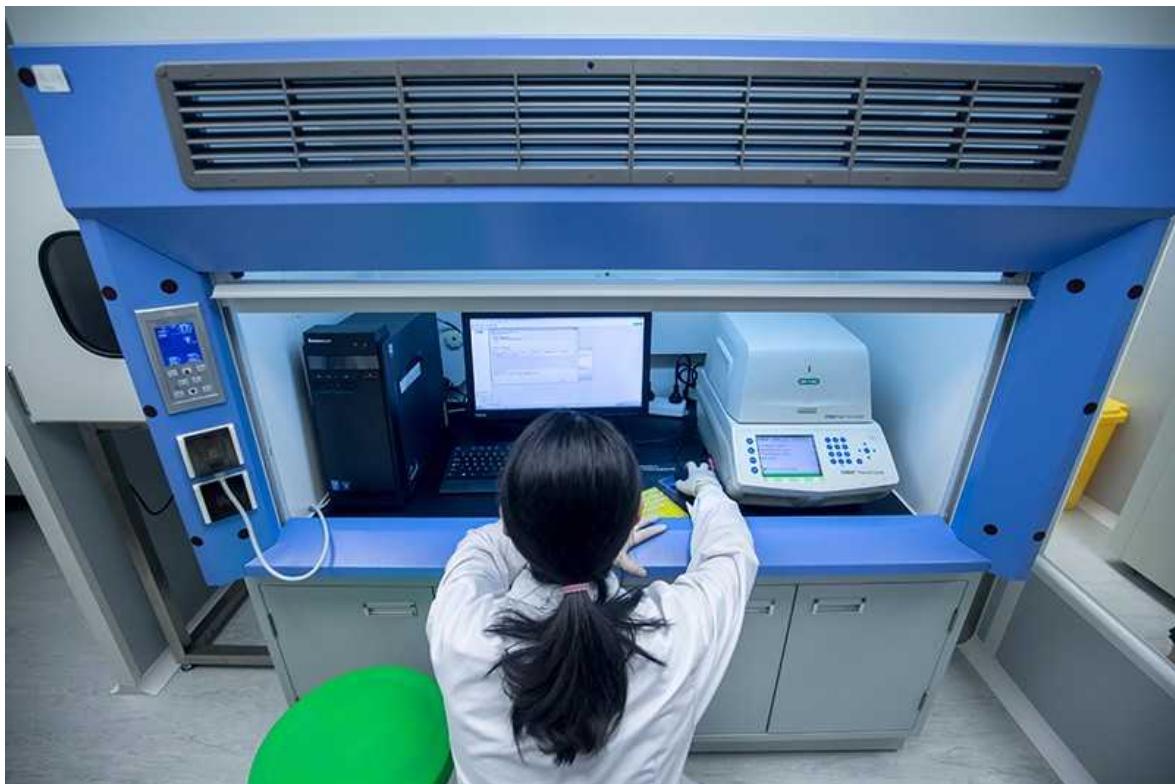
This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-022-01320-y>.

- NEWS
- 06 May 2022

China expands control over genetic data used in scientific research

Guidelines released this year are the latest regulations to protect China's genetic resources, but some scientists say they are making collaborations harder.

- [Smriti Mallapaty](#)



The sharing of genetic resources — including biological samples, such as organs, tissues and blood, that yield DNA, and the data gleaned from

sequencing them — is tightly controlled in China. Credit: Paul Yeung/Bloomberg via Getty

China is expanding its control over the use of genetic data from people in the country, including for scientific research. Researchers say this approach is making it harder for scientists there to collaborate with international peers.

A set of draft guidelines, released by China's Ministry of Science and Technology (MOST) in March, are the latest in a series of government orders on managing genetic resources — including biological samples, such as organs, tissues and blood, that yield DNA, and the data gleaned from sequencing them.

The guidelines provide detailed instructions on how to interpret and implement existing regulations. This includes 2019 and 2021 [laws that restrict](#) Chinese organizations from collecting certain types of genetic information and from sharing genetic resources with foreign groups.

The guidelines “paint a picture of increasing control”, says Jonathan Flint, a geneticist at the University of California, Los Angeles, who in 2015 published the results of a large collaboration analysing genetic data from people in China in *Nature*¹. “The regulatory landscape has shifted since then.”

Government officials say restrictions on the use of genetic data are designed to increase protection of this resource. They were developed in response to [companies exporting genetic data without permission](#), as well as other incidents, including the explosive revelation by Chinese researcher He Jiankui that he had created the first babies with edited genomes in 2018.

Many countries control how their citizens' DNA data can be shared, but most wealthy countries facilitate data sharing for research, says Arcadi Navarro, a geneticist at Pompeu Fabra University in Barcelona, Spain. In contrast, a history of scientific exploitation in low-income countries and those with vulnerable minority ethnic groups has resulted in some nations introducing strict data-sharing rules that are similar to China's, he says.

Negative effect

Shuhua Xu, a geneticist at Fudan University in Shanghai, China, says he supports the regulation of human genetic resources in principle, but thinks some of the requirements under the latest guidelines are too restrictive and will deter scientists from doing genetics work. These include the need to conduct a ‘security review’ to share data involving groups of more than 500 people, which is a relatively small number for studies on genetic diseases.

Under China’s rules, foreign organizations can collect and store genetic information from Chinese citizens only if they partner with a Chinese institution — and they need ministry permission to do so. Researchers say these requirements are making it hard for scientists in China to collaborate with international peers and to publish work in international journals that require data to be deposited in public repositories.

Xu says applying for permission from the MOST is complex and time-consuming. And it’s often unclear why permission is granted in some cases and not others. Last year, he published a study with US colleagues on the ancient origins of a gene found in Tibetan people, after being granted permission to share the data. But the MOST has previously rejected his requests to share data with international collaborators for studies on the genetic diversity and ancestry of ethnic groups, and he has heard of colleagues’ rejections. Rejections come with limited feedback on the denial, says Xu, who adds that many of his international collaborators have stopped asking for data.

The process might have improved since Xu’s last application in early 2021. The MOST website reports that, every month this year, it has approved several hundred applications to share data in international scientific collaborations — but it is not clear how many have been rejected.

Lost ties

Navarro worries that China’s increasing control will make it harder for Chinese scientists to deposit genetic data on publicly accessible repositories for researchers outside China to use. But so far, he hasn’t noticed a drop in the number of submissions from Chinese institutions in the European

Genome-phenome Archive (EGA) — a repository for genetic and phenotypic data that he is a group leader of.

Ultimately China's restrictive data-sharing policies will hurt local researchers the most because they will be isolated and “left out of the human genetics community”, says Flint. “It is a great shame.” In August 2021, he received a grant from the US National Institutes of Health to study the genetic causes of depression in people in South Korea, which he did not conduct in China because of the strict rules there.

Xu says he has been reluctant to join international genome consortiums, such as the COVID-19 Host Genetics Initiative, in the past few years because of the potential issues with sharing data. He thinks other Chinese researchers have probably made the same decision.

But the regulations might not be affecting all areas of population genetics. Choongwon Jeong, a population and evolutionary geneticist at Seoul National University, who studies ancient genomes, says the regulations haven't affected his partnerships with Chinese researchers. But he worries that China's tightening control could threaten this work in the future.

Nature **605**, 405 (2022)

doi: <https://doi.org/10.1038/d41586-022-01230-z>

References

1. CONVERGE Consortium. *Nature* **523**, 588–591 (2015).
2. Zhang, X., Witt, K. E., Bañuelos, M. M. & Huerta-Sánchez, E. *Proc. Natl Acad. Sci. USA* **118**, e2020803118 (2021).

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-022-01230-z>

- NEWS
- 06 May 2022
- Correction [09 May 2022](#)

This Arctic town wants to make renewable energy work at the top of the world

Partnering with a northern settlement in Greenland, researchers are designing wind and solar devices that can survive and thrive in extreme conditions.

- [Freda Kreier](#)



Qaanaaq, with its roughly 600 residents, is the northernmost town in Greenland. Credit: Mary Albert

For Toku Oshima, a hunter from Greenland, the quest to bring renewable energy to her hometown of Qaanaaq is not just a fight against climate change — it's a fight for cultural survival.

In 2015, Oshima journeyed from Qaanaaq — the northernmost town in the country — to a climate conference in Ilulissat, about 1,000 kilometres to the south. She had wanted to discuss how Earth's rapidly changing climate threatened traditional ways of life in Greenland, such as Indigenous forms of hunting and fishing. But she came away also thinking about how to tackle another existential crisis facing Qaanaaq: energy security.

Like many residents of the Arctic, the town's people struggle to pay for the fossil fuels they need to heat and power their homes, compounding many of the other pressures they face. In the past few years, hunters have found themselves unable to afford to feed their sledge dogs. And some residents are relocating to escape financial and mental depression. In the process, Oshima says, many Greenlanders are losing touch with their communities and culture.

At the conference, a friend told Oshima that she'd seen a presentation in which Mary Albert, a snow physicist at Dartmouth College in Hanover, New Hampshire, had discussed climate-change evidence that's preserved in ice cores. Albert had mentioned that transitioning to renewable energy sources could help to curb the climate crisis, and had shown a picture of her own home decked out with solar panels. Oshima approached Albert during a coffee break, thinking that the scientist might have ideas about how to cut energy costs in Qaanaaq.

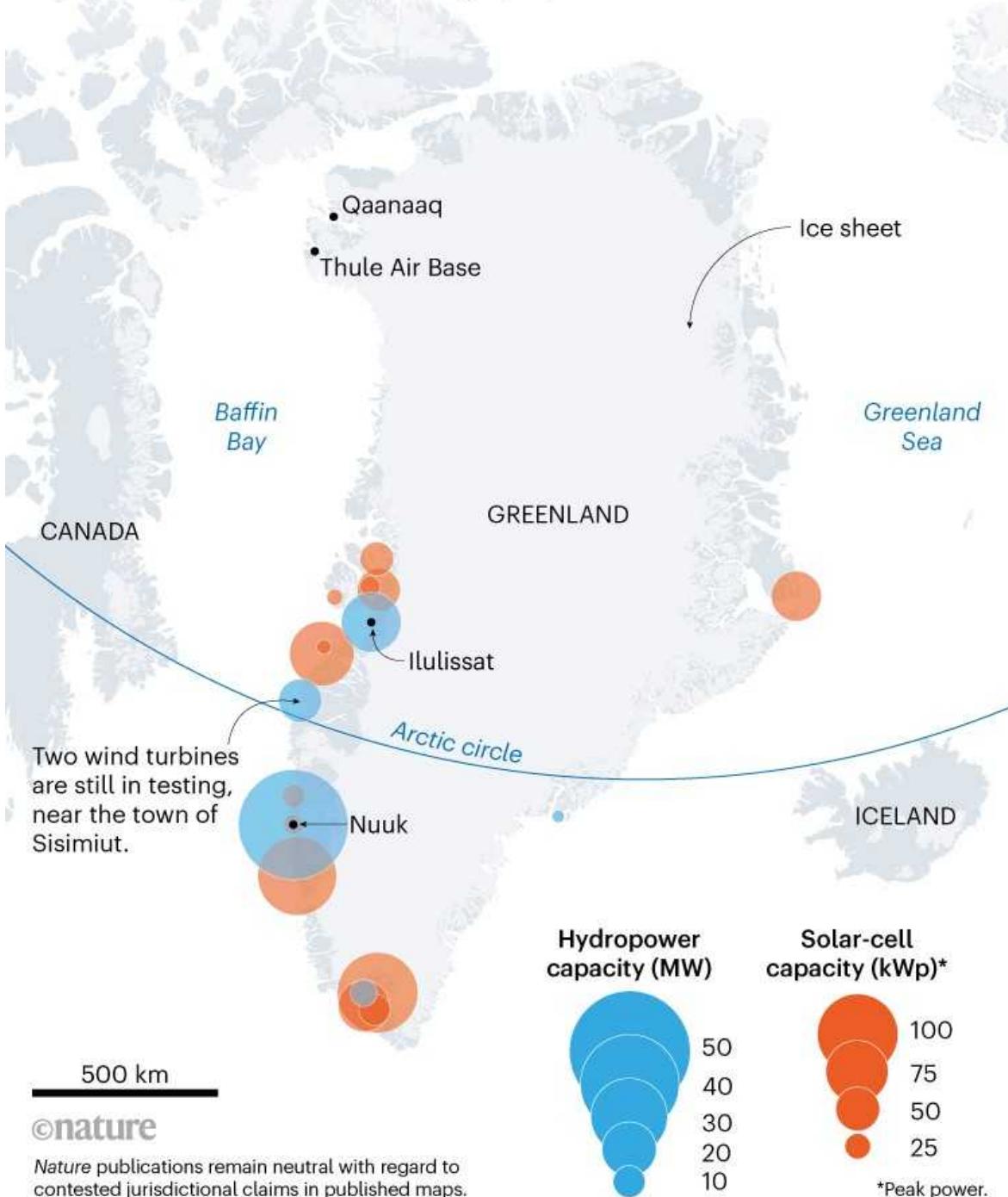
Albert was at first dubious that she could help. But as she opened her mouth to explain that renewables weren't her speciality, she thought to herself: "What a cop out. You have a PhD in engineering. Are you really going to tell these people there's no way you can help them?"

That marked the beginning of a collaboration between Albert and Oshima to help wean Qaanaaq off fossil fuels. In the past few years, Albert has recruited students at Dartmouth to work on a range of projects, including designing energy-efficient homes and testing the town's potential for solar energy. And this month, Albert is heading to Qaanaaq to test newly

developed solar and wind devices that she hopes will one day heat homes there.

GOING GREEN

Greenland already gets a majority of its electricity from renewable-energy sources such as hydroelectric dams and solar panels. But these sources have so far been concentrated in the populous southwestern part of the country, with more isolated communities such as Qaanaaq still largely dependent on fossil fuels.



Source: Niels Erik Hagelqvist/Nukissiorfuit

For both women, this project is about more than one town. Researchers and communities from Siberia to Canada's Northwest Territories aim to bring renewables to the Arctic. For scientists, that means designing technology that can work in places where the sun might not shine for months and temperatures stay decidedly below freezing for most of the year.

Although Greenland has made great strides in installing renewables, these changes have so far mostly benefited larger communities in the south of the country. Making cheap, accessible renewables work in Qaanaaq has the potential to "be good not only for this community, but for all Arctic areas", however remote, Oshima says.

At the top of the world

About 1,400 kilometres south of the North Pole, Qaanaaq was founded in 1953 after the US Air Force gave around 100 Inughuit people living in a town near the newly built Thule Airbase just a few days to pack up and head to Greenland's far north. Today, many of the town's approximately 600 residents help with the day-to-day running of Qaanaaq by, for instance, teaching at its school or operating its water plant. A dwindling number still make a living off the land, using dog sledges to travel out onto the thick sea ice in search of seals and halibut.

Among them is Oshima, who was born in Qaanaaq in 1975 to a family of hunters. Although she left to study electrical engineering in Greenland's capital, Nuuk, she eventually returned. She now hunts and runs a workshop where she cures skins and butchers meat. "I've never been an inside woman," she says. "Here, there are more chances to go out into nature."



Toku Oshima, a hunter, runs a workshop in Qaanaaq where she cures skins and butchers meat.Credit: Mary Albert

Like everyone else in the town, Oshima relies on electricity produced by the town's diesel generators. This is common in remote Arctic communities that were established after the Second World War, says Sherry Stout, a programme manager for Arctic strategies at the National Renewable Energy Laboratory in Golden, Colorado. Towns that emerged during this period often installed diesel generators because the fuel is less volatile than other types, and easier to transport over long distances, she says.

But getting diesel to such a remote location isn't easy. Like many Arctic communities, Qaanaaq is not connected to the rest of the world through roads. Almost everything that the town doesn't itself produce — including its supply of diesel — arrives on an icebreaker ship. Two deliveries take place each year, both of them in the Arctic's late summer, when sea ice is at its lowest.

“Most communities only have one shot at getting all the diesel they need for the year,” says Martha Lenio, a renewable-energy specialist in the conservation group WWF’s office in Iqaluit, Canada. Increasingly erratic weather and sea ice sometimes means that the ships the communities rely on never come in. In 2019, unusual ice conditions prevented barges from making their annual diesel delivery to several hamlets in Canada’s Northwest Territories; in the end, the shipment had to be flown in.

Diesel despair

These logistics explain why the cost of fuel is so high in Arctic communities, Stout says. Making electricity from fossil fuels in the United States costs about 14 cents per kilowatt hour, on average. But in northern parts of Alaska, that price jumps to between 50 cents and US\$1 per kilowatt hour.

Residents of Qaanaaq spend a considerable amount on fuel; sub-zero temperatures mean houses must be heated nearly continuously. Complicating the situation is that many of Qaanaaq’s houses are Danish-style — some of them dating back to 1953 — and these are not efficient at holding in heat.

To bring costs down, Greenland’s government heavily subsidizes fossil fuels, says Niels Erik Hagelqvist, a renewable-energy adviser at Nukissiorfiit, the country’s state-owned energy company in Nuuk. But even subsidized diesel — whose price equates to around 24 cents per kilowatt hour in US currency — is a significant expense for Qaanaaq’s residents, especially hunters who make little money selling their excess catch, Albert says.

“It’s not a simple life if you want to stay here,” Oshima says. “It’s very expensive.”

Which is one reason Oshima would like to bring renewables to Qaanaaq. But finding the funds to make this transition is tricky. Greenland announced in 2017 that it planned to switch to renewables as much as possible by 2030. But all technologies have to be shipped in, says Hagelqvist, making these

projects extremely costly. And with Greenland still dependent on funds from Denmark — it was a colony of Denmark but won home rule in the late 1970s — “our ambitions are limited by a lack of money”, he says.



Snow physicist Mary Albert visited Qaanaaq with her students in 2019. She is returning this month to test renewable devices. From left: Joshua Elliot (student at the time), Lene Kielsen Holm (key research partner in Greenland who passed away in 2021), Kim Petersen (Oshima’s husband), Toku Oshima, Albert and Hunter Snyder (now a postdoc). Credit: Hunter T. Snyder

This, in part, explains why Greenland’s 5 hydroelectric dams and 13 solar panel farms are concentrated in the more populous southwestern part of the country, where they can benefit the largest number of people (see ‘Going green’). Small, northerly communities such as Qaanaaq typically need to raise their own money — or apply for government grants — if they want to make renewable energy a reality. And this can be risky, says Robert Cooke, a technology officer at the Canadian High Arctic Research Station in Whitehorse, Canada.

For instance, in 2007, the small Alaskan community of Nikolski built a wind turbine using a government grant, but the turbine sat unused for years because technicians struggled to connect it to the town’s power-plant control panels. Most renewables were not built to handle Arctic conditions, Cooke says, and if a system is damaged, “it can be weeks before somebody can fly in and fix it”.

Prototypes on the way

Albert and her students hope to get around these problems by building technologies that Qaanaaq's carpenter — Oshima's husband — can easily install and fix.

Two of the group's devices will be tested in Qaanaaq this month. One of these prototypes, designed by Albert's student Simon Oster, uses wind energy to generate heat. Other renewable-energy projects in the Arctic, such as the Chaninik Wind Group in Alaska, do this already, using electricity produced by wind turbines. But Albert says that Oster's device is different because it generates heat directly from wind. Similar designs have been suggested before, Oster says, but none has reached the prototype stage. His own design is "top secret" for now, so he can't discuss it.

Albert hopes that the wind device can be used in the dark of winter to reduce heating costs. And in the summer, homes could generate heat using the second prototype. This one, designed by Dartmouth student Tucker Oddleifson, is a type of window that traps solar heat. The idea is not new, but Oddleifson says that these particular windows are simpler than other designs and are built to make the most of the Arctic summer's 24 hours of sunlight. The team hopes that, eventually, "these small-scale devices could be used anywhere", Albert says.

Oshima and the researchers are not under any illusions that their project will solve all the town's challenges. But the devices could go a long way to helping Qaanaaq's residents have an easier life and continue to "live in northwestern Greenland, where their people have lived for thousands of years", Albert says. "They want to be warm and comfortable — and they want to be able to afford it."

Nature **605**, 406-407 (2022)

doi: <https://doi.org/10.1038/d41586-022-01189-x>

Updates & Corrections

- **Correction 09 May 2022:** An earlier version of this story referred to Dartmouth University in New Hampshire. It is Dartmouth College.
-

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-022-01189-x>

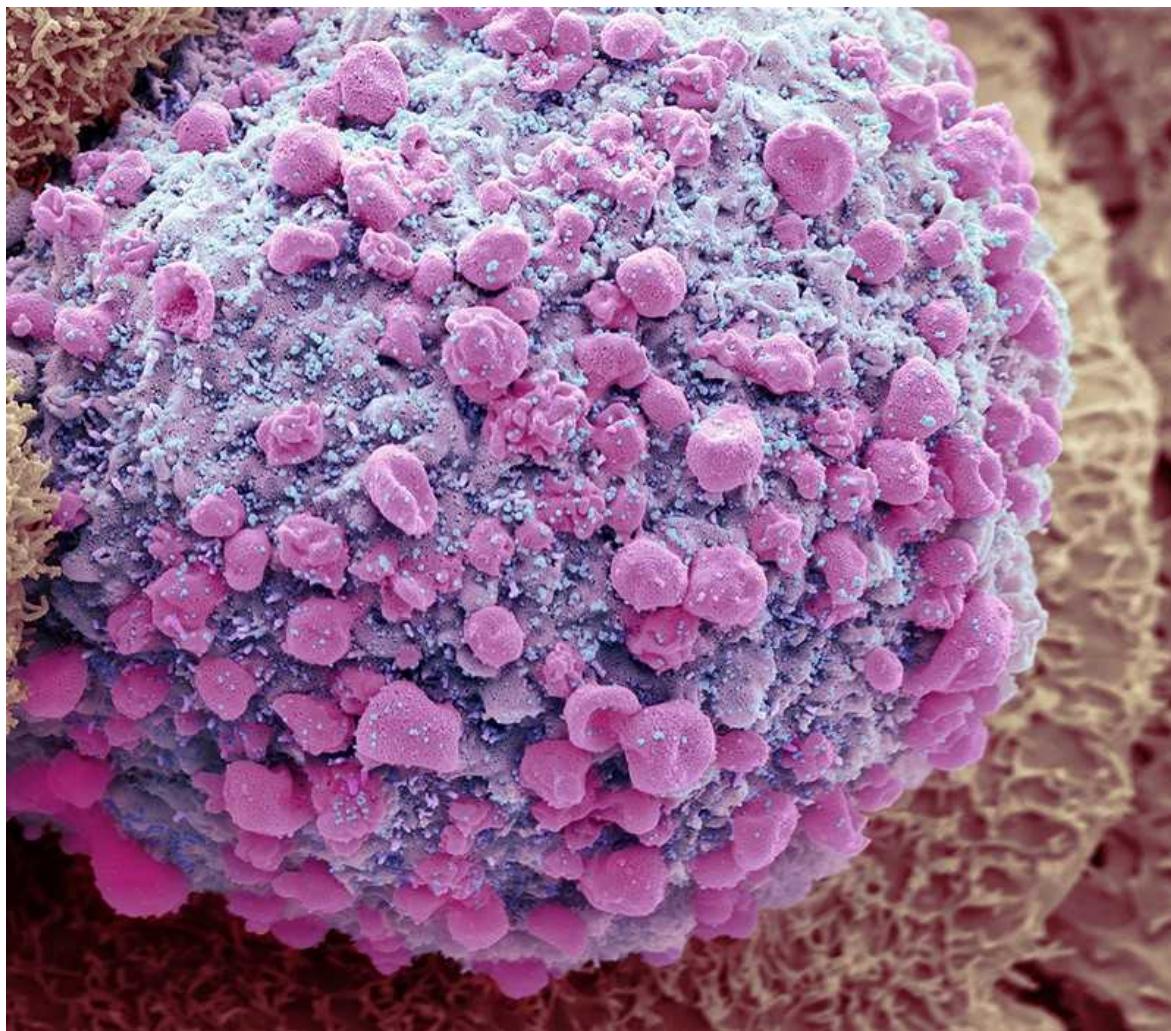
| [Section menu](#) | [Main menu](#) |

- NEWS
- 11 May 2022

Coronavirus ‘ghosts’ found lingering in the gut

Scientists are studying whether long COVID could be linked to viral fragments found in the body months after initial infection.

- [Heidi Ledford](#)



Particles of SARS-CoV-2 (blue; artificially coloured) bud from a dying intestinal cell. Credit: Steve Gschmeissner/SPL

In the chaos of the first months of the coronavirus pandemic, oncologist and geneticist Ami Bhatt was intrigued by widespread reports of vomiting and diarrhoea in people infected with SARS-CoV-2. “At that time, this was thought to be a respiratory virus,” she says. Bhatt and her colleagues, curious about a possible link between the virus and the gastrointestinal symptoms, began to collect stool samples from people with COVID-19.

Thousands of miles away from Bhatt’s laboratory at Stanford Medicine in California, gastroenterology internist Timon Adolph was puzzled by accounts of gut symptoms in infected people. Adolph and his colleagues at the Medical University of Innsbruck in Austria started to assemble specimens, too — gastrointestinal-tissue biopsies.

Two years into the pandemic, the scientists’ foresight has paid off: both teams have recently published results^{1,2} suggesting that pieces of SARS-CoV-2 can linger in the gut for months after an initial infection. The findings add to a growing pool of evidence supporting the hypothesis that persistent bits of virus — coronavirus “ghosts”, Bhatt has called them — could contribute to the mysterious condition called long COVID.

Even so, Bhatt both urges scientists to keep an open mind and cautions that researchers have not yet nailed down a link between persistent viral fragments and long COVID. “Additional studies still need to be done — and they’re not easy,” she says.

Long COVID is often defined as symptoms that linger beyond 12 weeks after an acute infection. More than 200 symptoms have been associated with the disorder, which ranges in severity from mild to debilitating. [Theories about its origins vary](#), and include harmful immune responses, tiny blood clots and lingering viral reservoirs in the body. Many researchers think that a mix of these factors contributes to the global burden of disease.

An early hint that the coronavirus might persist in the body came in work³ published in 2021 by gastroenterologist Saurabh Mehandru at the Icahn School of Medicine at Mount Sinai in New York City and his colleagues. By

then, it was clear that cells lining the gut display the protein that the virus uses to enter cells. This allows SARS-CoV-2 to infect the gut.

Mehandru and his team found viral nucleic acids and proteins in gastrointestinal tissue collected from people who'd been diagnosed with COVID-19 an average of four months earlier. The researchers also studied participants' memory B cells, which are pivotal players in the immune system. The team found that antibodies produced by these B cells were continuing to evolve, suggesting that, at six months after the initial infection, the cells were still responding to molecules made by SARS-CoV-2.

Inspired by this work, Bhatt and her colleagues found that a few people continued to shed viral RNA into their stool seven months after an initial mild or moderate SARS-CoV-2 infection, well after their respiratory symptoms had ended¹.

Virus goes for the gut

Adolph says the 2021 paper inspired his team to look at their biopsy samples for signs of coronavirus. They found that 32 of 46 study participants who had had mild COVID-19 showed evidence of viral molecules in their gut seven months after acute infection. About two-thirds of those 32 people had long-COVID symptoms.

But all of the participants in this study had inflammatory bowel disease, an autoimmune disorder, and Adolph cautions that his data do not establish that there is active virus in these people, or that the viral material is causing long COVID.

In the meantime, more studies have suggested lingering viral reservoirs beyond the gut. Another team of researchers has studied tissue collected from autopsies of 44 people who had been diagnosed with COVID-19 and found evidence of viral RNA in many sites, including the heart, eyes and brain⁴. Viral RNA and proteins were detected up to 230 days after infection. The study has not yet been peer reviewed.

Viral hideouts

Nearly all of the people in that sample had had severe COVID-19, but a separate study of two people who had had mild COVID-19 followed by long COVID symptoms found viral RNA in the appendix and the breast⁵.

Pathologist Joe Yeong at the Institute of Molecular and Cell Biology at the Agency for Science, Technology, and Research in Singapore, who is a co-author of the report, which has not been peer reviewed, speculates that the virus might infiltrate and hide out in immune cells called macrophages, which can be found in a variety of the body's tissues.

All of these studies support the possibility that long-term viral reservoirs contribute to long COVID, but researchers will need to do more work to conclusively show a link, says Mehandru. They will need to document that the coronavirus is evolving in people who are not immunocompromised, and they will need to link such evolution to long COVID symptoms. "Right now there is anecdotal evidence, but there are a lot of unknowns," Mehandru says

Bhatt is hopeful that samples will become available to test the viral-reservoir hypothesis. The US National Institute of Health, for example, is running a large study called RECOVER, which aims to tackle the causes of long COVID and will collect biopsies from the lower intestines of some participants.

But Sheng says he does not need to wait for a billion-dollar study to get more samples: an organization of people with long COVID has contacted him and offered to send samples from members who have had biopsies for various reasons, such as a cancer diagnosis, after their infections. "It's really random, the tissue can come from everywhere," he says. "But they don't want to wait."

Nature **605**, 408-409 (2022)

doi: <https://doi.org/10.1038/d41586-022-01280-3>

References

1. Natarajan, A. *et al. Med* <https://doi.org/10.1016/j.medj.2022.04.001> (2022).

2. Zollner, A. *et al.* *Gasteroenterology*
<https://doi.org/10.1053/j.gastro.2022.04.037> (2022).
3. Gaebler, C. *et al.* *Nature* **591**, 639–644 (2021).
4. Chertow, D. *et al.* Preprint at Research Square
<https://doi.org/10.21203/rs.3.rs-1139035/v1> (2021).
5. Goh, D. *et al.* Preprint at Research Square
<https://doi.org/10.21203/rs.3.rs-1379777/v1> (2022).

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-022-01280-3>

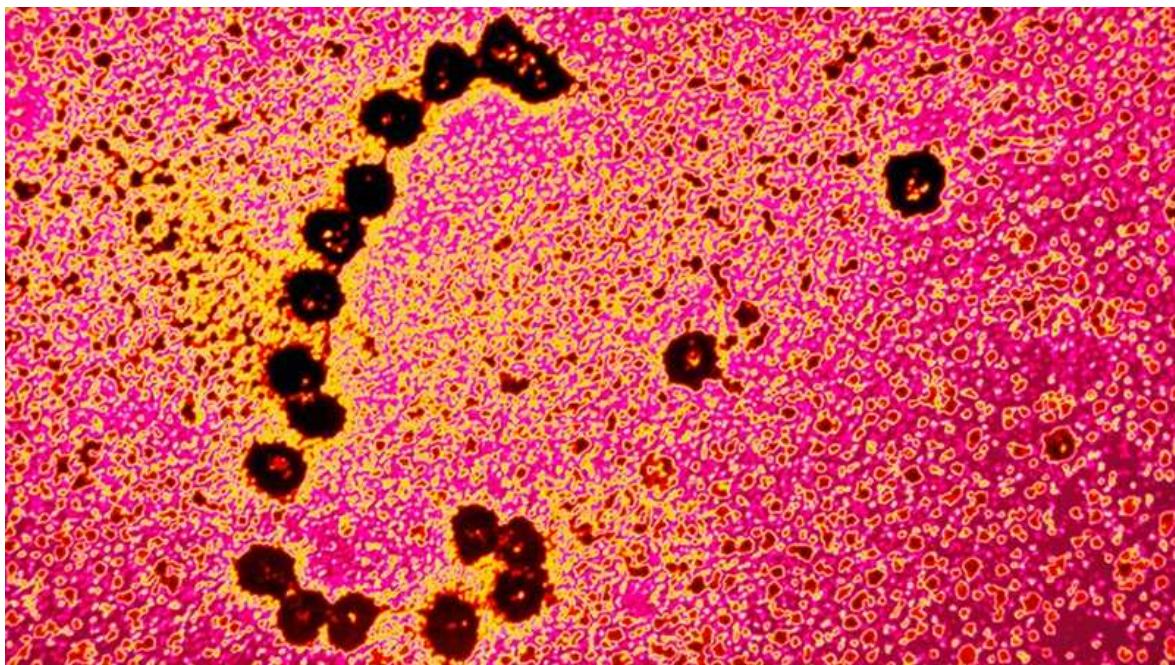
| [Section menu](#) | [Main menu](#) |

- NEWS
- 11 May 2022

Origin of life theory involving RNA–protein hybrid gets new support

Structure that links amino acids suggests that early organisms could have been based on an RNA–protein mix.

- [Davide Castelvecchi](#)



Carell and colleagues were inspired by ribosomes — shown here translating a strand of RNA.Credit: Omikron/Science Photo Library

Chemists say they have solved a crucial problem in a theory of life's beginnings, by demonstrating that RNA molecules can link short chains of

amino acids together.

The findings, published on 11 May in *Nature*¹, support a variation on the '[RNA world' hypothesis](#)', which proposes that before the evolution of DNA and the proteins it encodes, the first organisms were based on strands of RNA, a molecule that can both store genetic information — as sequences of the nucleosides A, C, G and U — and act as a catalyst for chemical reactions.

The discovery “opens up vast and fundamentally new avenues of pursuit for early chemical evolution”, says Bill Martin, who studies molecular evolution at Heinrich Heine University Düsseldorf in Germany.

In an RNA world, the standard theory says, life could have existed as complex proto-RNA strands that were able to both copy themselves and compete with other strands. Later, these ‘RNA enzymes’ could have evolved the ability to build proteins and ultimately to transfer their genetic information into more-stable DNA. Exactly how this could happen was an open question, partly because catalysts made of RNA alone are much less efficient than the protein-based enzymes found in all living cells today. “Although [RNA] catalysts were discovered, their catalytic power is lousy,” says Thomas Carell, an organic chemist at Ludwig Maximilian University of Munich in Germany.

RNA ribosome

While investigating this conundrum, Carell and his collaborators were inspired by the part that RNA plays in how all modern organisms build proteins: a strand of RNA encoding a gene (typically copied from a sequence of DNA bases) passes through a large molecular machine called a ribosome, which builds the corresponding protein one amino acid at a time.

Unlike most enzymes, the ribosome itself is made of not only proteins, but also segments of RNA — and these have an important role in synthesizing proteins. Moreover, the ribosome contains modified versions of the standard RNA nucleosides A, C, G, and U. These exotic nucleosides have long been seen as possible vestiges of a primordial broth.

Carell's team built a synthetic RNA molecule that included two such modified nucleosides by joining two pieces of RNA commonly found in living cells. At the first of the exotic sites, the synthetic molecule could bind to an amino acid, which then moved sideways to bind with the second exotic nucleoside adjacent to it. The team then separated their original RNA strands and brought in a fresh one, carrying its own amino acid. This was in the correct position to form a strong covalent bond with the amino acid previously attached to the second strand. The process continued step by step, growing a short chain of amino acids — a mini-protein called a peptide — that grew attached to the RNA. The formation of bonds between amino acids requires energy, which the researchers provided by priming the amino acids with various reactants in the solution.

"This is a very exciting finding," says Martin, "not only because it maps out a new route to RNA-based peptide formation, but because it also uncovers new evolutionary significance to the naturally occurring modified bases of RNA." The results point to an important part played by RNA at the origins of life, but without requiring RNA alone to self-replicate, Martin adds.

Loren Williams, a biophysical chemist at the Georgia Institute of Technology in Atlanta, agrees. "If the origins of RNA and the origins of protein are linked, and their emergence is not independent, then the math shifts radically in favour of an RNA–protein world and away from an RNA world," he says.

To show that this is a plausible origin of life, scientists must complete several further steps. The peptides that form on the team's RNA are composed of a random sequence of amino acids, rather than one determined by information stored in the RNA. Carell says that larger RNA structures could have sections that fold into shapes that 'recognize' specific amino acids at specific sites, producing a well-determined structure. And some of these complex RNA–peptide hybrids could have catalytic properties, and be subject to evolutionary pressure to become more efficient. "If the molecule can replicate, you have something like a mini organism," says Carell.

Nature **605**, 409 (2022)

doi: <https://doi.org/10.1038/d41586-022-01303-z>

References

1. Müller, F. *et al.* *Nature* <https://doi.org/10.1038/s41586-022-04676-3> (2022).

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-022-01303-z>

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

- NEWS FEATURE
- 18 May 2022

The pandemic's true health cost: how much of our lives has COVID stolen?

Researchers are trying to calculate how many years have been lost to disability and death.

- [Holly Else](#)



Millions of people are living with lasting effects of COVID-19. Credit: Mark Mirko/Hartford Courant/Tribune News Service/Getty

How do you count the cost of a pandemic? COVID-19 has [killed an estimated 15 million people](#) since it emerged at the end of 2019, but its impact on health reaches much further. For hundreds of millions of people around the world, infection with the coronavirus SARS-CoV-2 has brought a range of problems, from the acute effects of the illness to the [lasting symptoms known as long COVID](#).

Working out the size of that health burden is challenging, but important — governments use such figures to plan how to spend health-care budgets. So researchers are starting to tally the overall health impacts and trying to draw lessons from any patterns. They're hoping, for example, to discern how different populations are affected and to provide evidence about the effects of vaccine roll-outs and new variants of the virus.

Even without a pandemic, there is no easy way to tally all the effects of various health conditions: good data can be hard to come by and decisions on how to measure burdens are inherently subjective. “There are a whole lot of social value choices where there isn’t hard science,” says Theo Vos, an epidemiologist at the Institute for Health Metrics and Evaluation (IHME) at the University of Washington in Seattle, a research centre that aims to categorize the global health burden of diseases. “How do you value a year with asthma, a year without a leg, a year with depression?”

Those calculations are even harder when scientists are grappling with a new virus and a poorly characterized disease.

Research groups are exploring a number of ways to calculate the burden of COVID-19, and many are starting to report their results. Early data suggest that the impact is significant and varies by country. One study found that COVID-19 took a heavy toll across 16 European countries, but that the impacts on different nations varied owing to factors ranging from the population’s age structure to political responses to the pandemic¹ (see ‘A heavy burden’).

Estimates produced by national teams provide more detail. In Scotland², COVID-19 was second only to ischaemic heart disease in terms of the impact it had on the population’s health in 2020. In the Netherlands³ that

year, the burden was 16 times that of a typical influenza season, according to a preprint published last November.

A HEAVY BURDEN

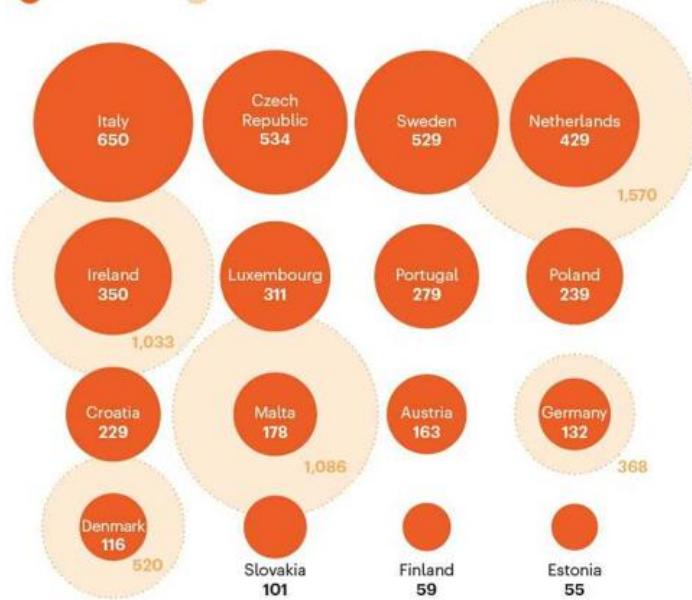
Researchers are starting to analyse how many years of life have been lost to death and ill health from COVID-19, measured in units called disability-adjusted life years (DALYs).

The toll of COVID-19 in Europe

An analysis of 16 European countries using data from sources such as the World Health Organization suggests that the toll can vary widely depending on a population's age structure or, perhaps, a government's response to the pandemic. National estimates, using more detailed data, tend to yield larger burden estimates.

COVID-19 DALYs per 100,000 people (data for 2020):

● European study ● Country-level data (where available)



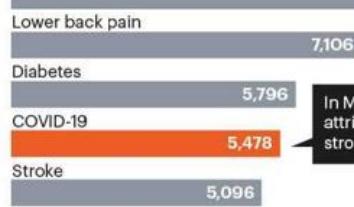
How COVID-19 stacks up

In countries for which data are available, COVID-19 tends to be one of the most burdensome diseases.

Malta

Ischaemic heart disease

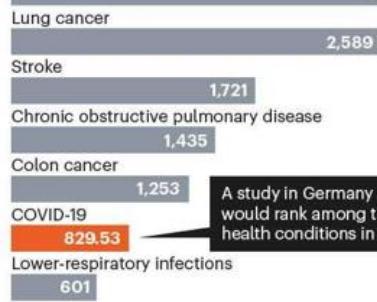
13,594 DALYs



Germany

Ischaemic heart disease

4,684 years of life lost to death (YLL)*



*Data are for 2020 for COVID-19 and for 2017 for other conditions.

Sources: European study: Ref 1; country data: ref 5, S. Monteiro Pires, A. Rommel *et al.* *Dtsch Arztebl Int.* **118**, 145–151 (2021)

With the pandemic still raging across many parts of the world, it's too early to calculate the full toll. But some researchers think it has helped to change how they calculate the health effects of diseases. "The pandemic has strengthened collaborations in the field of disease burden," says Sara Monteiro Pires, an epidemiologist at the Technical University of Denmark. Researchers are now harmonizing the processes they use to estimate disease burden, and tailoring the models to the data available in each location. They hope this will make the results more precise.

Adding up the effects

The UK National Health Service lists a dozen COVID-19 symptoms for adults, from loss of smell to a high temperature. Even people who have a relatively mild case and ride out the symptoms at home can see lasting health effects, such as fatigue or shortness of breath. If symptoms continue beyond a couple of months, people can be diagnosed with an illness widely known as long COVID.

To quantify how a disease affects an entire population, scientists combine data on individual experiences. These include the number of people infected, the number who had certain symptoms, the length of illnesses, how many needed hospital treatment or died, and patients' ages, among other things. They then use them to work out how many years of life have been lost to the disease and how many years are lived with disabling symptoms.

Researchers can use the average life expectancies in a country to work out how many years of life have been lost owing to premature death. Losses due to disability are harder to calculate, however. To quantify those, researchers use data about the number of people affected by a certain illness, the length of time they have it and a value for the illness known as a disability weight. The IHME's Global Burden of Disease group maintains a standardized list of disability weights; the latest version available, published in 2019, gives a mild earache a disability weight of 0.013 and severe multiple sclerosis 0.719 (a weight of 0 is perfect health; a weight of 1 is death).

Currently, there is no standardized disability weight for COVID-19. Instead, researchers use the disability weights associated with other infectious diseases and similar health conditions.

Totalling the years of life lost due to illness, disability or premature death gives an estimate of the burden in a unit known as disability-adjusted life years, or DALYs. It is the keystone of research into the burden of disease.

The data that go into DALYs come from a variety of sources. Many are routinely collected by national health authorities. For COVID-19, some data have been gathered through targeted surveillance efforts such as the REACT study (Real-time Assessment of Community Transmission), a massive sampling exercise that began in 2020 and has chronicled how SARS-CoV-2 is moving through England and what symptoms people are experiencing.

Data from the REACT study suggest that COVID-19's health effects can linger. A preprint posted on the medRxiv server last July suggested that 19% of the English population had had COVID-19, and that around one-third of those — more than 2 million adults — had experienced one or more symptoms for at least 12 weeks⁴. “That’s 6% of the population,” says Paul Elliot, an epidemiologist at Imperial College London, who leads the REACT study and co-authored the study.

Max Taquet, a clinical researcher and engineer at the University of Oxford, UK, who uses data from medical records to understand the neurological and psychiatric consequences of COVID-19, says that estimating the health effects of long COVID is difficult and that the numbers are startling. “Many of us were surprised of the scale of the problem,” he says, “but we do see this [post-infection syndrome] with other viral infections”. With COVID-19, scientists are monitoring the consequences in real time. “It’s great that we are finally paying attention to it.”

There is no guarantee that these data sources will stick around, however. The UK government announced in March that it would be stopping funding for some branches of the REACT study and for another surveillance effort.

Big burden

Early results on the health lost to COVID-19 are trickling in. “Overall, the impact of COVID-19 has been dramatically high worldwide,” says Gianfranco Politano, a bioinformatician at the Polytechnic University of Turin in Italy, who was involved in the study of 16 European countries.

The European research suggests that Slovakia probably had a lower burden than other countries because the government acted quickly to lock down and people complied. By contrast, the burden was higher in Sweden, where the government took a “herd immunity” approach and allowed the virus to spread largely unchecked.



A woman who has had COVID-19 attends a physiotherapy session in Madrid. Credit: Pierre-Phillipe Marcou/AFP/Getty

Individual-country analyses also reveal big differences in the health burden of COVID-19. Research from Malta reveals that between March 2020 and March 2021, COVID-19 became the fourth leading cause of disability, ranking after ischaemic heart disease, lower back pain and diabetes⁵. In India, it ranked much further down the list: using 2019 data as a guide, it

would have accounted for 3% of the total health burden — putting it outside the top 10 and rating it as less of a burden than ischaemic heart disease, nutritional deficiencies and chronic respiratory diseases⁶. The authors acknowledge, however, that COVID-19 cases might be under-reported in India, which would affect the rate of DALYs.

Each project sources its data slightly differently, which can add to the variation in DALY estimations. The research group that estimated DALYs for 16 European countries, for example, used aggregated data from the European Centre for Disease Prevention and Control (ECDC), the World Health Organization (WHO) and the World Bank Group; many of the national studies used more-detailed country-specific data. As a consequence, DALY estimates for the same country vary in different hands. Using ECDC, WHO and World Bank data for Denmark, for example, gives a figure of 116 DALYs per 100,000 people¹, whereas Monteiro Pires' group used data from Denmark's health systems to come up with a figure closer to 520 (see go.nature.com/3m6nsrj).

Several of the studies of individual European countries have been supported by the European Burden of Disease Network, a project launched in 2019 to improve how the burden of disease is calculated and understood. The network of epidemiologists and public-health researchers from 53 countries worldwide quickly realized that it should be documenting the public-health burden of the nascent pandemic virus, and began to develop a consensus protocol, including a specific model for the disease progression of COVID-19 from infection to recovery or death. “From that moment on, many countries have been using that protocol. We never imagined it would happen so quickly,” says Monteiro Pires, who heads the network’s infectious-disease working group. Network researchers have now completed burden estimates for Malta, Denmark, the Netherlands, Scotland, Ireland and Germany, with more expected to appear in the months ahead.

One important job for the network was to align the definitions used in the data sets so that the burdens of disease could be compared across countries. But it is still too soon to draw any major conclusions from the work, says Monteiro Pires.

There is not yet an estimate of the global health toll from COVID-19, but the IHME has been churning out figures for a catalogue of other diseases since the 1990s. In early 2020, when it became clear a pandemic was under way, the institute already had the machinery in place to help it understand the wider health effects of SARS-CoV-2 and got to work on adding COVID-19 to the catalogue. Around 100 staff members were diverted to the effort. Their data are currently being considered for publication.

In contrast to many other calculations, the data include estimates of the burden of long COVID. Vos has presented these unpublished data to US authorities to help them get a handle on how the lingering symptoms could affect people's ability to work. The findings suggest that in 2020 and 2021, an estimated 4.6 million people in the United States had symptoms that persisted for at least three months. The group's definition of long COVID revolves around three clusters of symptoms, centring on fatigue, cognitive problems and ongoing respiratory issues. More than 85% of these cases came as a result of a bout of COVID-19 that did not require hospital treatment.

"It's a sizeable problem, these are people who are pretty severely disabled," says Vos.

The team's modelling suggests that around 5% of women and 2% of men who had a mild case of COVID-19 still had symptoms 6 months after the acute phase of the illness ended. For those treated in hospital, it was 26% of women and 15% of men, rising to 42% and 27%, respectively, if the patient spent time in the intensive care unit.

Vos' team found that people with long COVID had an average disability weight of 0.21 — equivalent to complete hearing loss or severe traumatic brain injury. "Hopefully this will trigger awareness with treating physicians that this is not trivial and it does exist," Vos adds.

Data gaps

One big problem for researchers attempting to estimate the burden of COVID-19 is the coverage of data. Some countries, such those in the Pacific

Islands, record so few cases that the data are not statistically sound. And many countries in sub-Saharan Africa, among other regions, lack the ability to track excess deaths due to COVID because of inadequate registration systems.

The IHME group get around this by using data from neighbouring countries to generate country-specific estimates. But ultimately, accurate calculations will require the collection of more detailed data. “People don’t automatically think that improving information systems is a priority in a pandemic,” says Andrew Briggs, a health economist at the London School of Hygiene & Tropical Medicine, “but in terms of preparedness we should be.” He and his colleague Anna Vassall recently predicted that as much as 30% of the health burden of COVID-19 could be down to disability⁷, not death.

The second data blind spot is long COVID. So far, only a few research groups outside the IHME have included such data in their estimates. Others think that without good information on long COVID, calculating the burden of the disease is premature.

Some national estimates — such as those for Scotland², Malta⁵ and Ireland⁸ — include limited long-COVID data in their analyses, but acknowledge the uncertainties. Grant Wyper, who works on the burden of disease for Public Health Scotland, helped to put together these estimates and says that the data on long COVID were sparse and that the condition was often defined in different ways — combining data from people who had just one symptom, such loss of sense of smell, with those from people who had several symptoms, which would have a more severe impact on quality of life.

Because so little was known when they developed the initial disease model, Wyper and his group used a general disability weight for the health effects seen after an infection. They are now working to refine the disability weighting for long COVID to make it more accurate, he says.

For its estimates of the burden of long COVID, the IHME sought out ongoing cohort studies that were logging symptoms and, in some instances, assessments of general health before COVID-19 developed. Its model pulls together data from 10 cohorts worldwide and includes more than 5,000

people treated in the community or hospital, as well as data from medical records and published studies.

But the estimates rely on the assumption that people who do not have symptoms during the acute phase do not develop long COVID. Taquet says that it is not yet clear that this is the case. “There is no reason to believe that someone with no symptoms at the time of the acute infection won’t go on to develop symptoms of long COVID later on,” he adds. His team has found that 2 in every 5 people with long COVID symptoms 3–6 months after infection did not report symptoms in the first 3 months⁹.

Some groups might be disproportionately affected by COVID-19. Briggs and Vassall stress that the data should be collected in a way that is sensitive to that, and broken down by age, socioeconomic and ethnic group. “As we move to an endemic situation, we have got to be more concerned by equity,” he says. For its part, the European Burden of Disease Network is hoping to look at how social inequality affects health burden in the future.

Measuring DALYs takes time — often the analyses are done only once a year. That means that some key questions about the burden of COVID-19 — such as how vaccines have affected illness rates and severity — won’t be answered for a while. The fact that COVID-19 has been around for only a couple of years means that scientists don’t have enough data to make accurate forecasts, says Maria Gianino, an economist at the University of Turin, who worked on the study of 16 European countries.

Despite the challenges, Monteiro Pires thinks that the future for disease-burden studies is bright. More funding is coming their way, she says. “It is more acknowledged that this is an important tool for public health”.

Nature **605**, 410–413 (2022)

doi: <https://doi.org/10.1038/d41586-022-01341-7>

References

1. Gianino, M. M. *et al. Eur. Rev. Med. Pharmacol. Sci.* **25**, 5529–5541 (2021).

2. Wyper, G. M. A. *et al.* *Arch. Public Health* **80**, 105 (2022).
3. McDonald, S. A. *et al.* Preprint at Research Square
<https://doi.org/10.21203/rs.3.rs-1026794/v1> (2021).
4. Whitaker, M. *et al.* Preprint at medRxiv
<https://doi.org/10.1101/2021.06.28.21259452> (2021).
5. Cuschieri, S., Calleja, N., Devleesschauwer, B. & Wyper, G.M.A. *BMC Public Health* **21**, 1827 (2021).
6. Singh, B. B. *et al.* *Sci. Rep.* **12**, 2454 (2022).
7. Briggs, A. & Vassall, A. *Nature* **593**, 502–505 (2021).
8. Moran, D. *et al.* Preprint at medRxiv
<https://doi.org/10.1101/2021.12.29.21268120> (2022).
9. Taquet, M. *et al.* *PLoS Med.* **18**, e1003773 (2021).

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-022-01341-7>

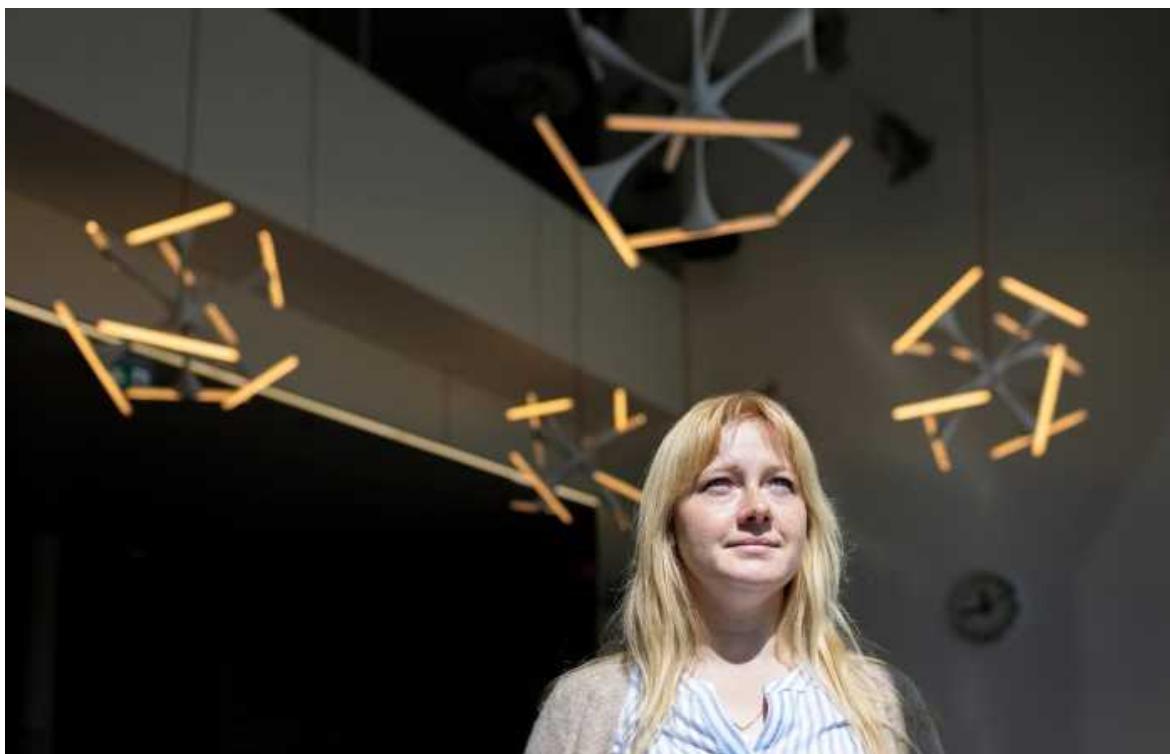
| [Section menu](#) | [Main menu](#) |

- NEWS FEATURE
- 06 May 2022
- Correction [09 May 2022](#)

How three Ukrainian scientists are surviving Russia's brutal war

Helping, healing and fighting: researchers have become refugees, soldiers and activists in the face of a horrifying conflict.

- [Nisha Gaind](#)



Ukrainian plasma physicist Olena Prysiazhna is now in the Netherlands and is talking to Dutch scientists about continuing her research. Credit: Ilvy Njiokiktjien/VII Photo for *Nature*

Olena Prysiazhna fled Russia's invasion twice. On 25 February, the 35-year-old plasma physicist raced out of Kyiv to her home village 80 kilometres away, hoping to escape the coming attacks on Ukraine's capital city. Two weeks later, Russian shells began raining down on the previously peaceful village. A rocket exploded in her neighbour's back garden.

"It broke our windows, doors, roof, but no one was hurt, thank God," says Prysiazhna. "After that, there were several attacks and we had to act."

Prysiazhna knew it was time to get out of Ukraine. With her sister Oksana, her mother and her German shepherd puppy Tokay, she set out to leave. After several days traversing the country, they walked across the border to Poland with no clear plan as to where they were going.

In the 10 weeks since [the Russian invasion](#), an estimated 3,100 civilians have died in Ukraine and more than 5 million Ukrainians have fled the country — creating [Europe's biggest refugee crisis in a generation](#). The war will indelibly alter the lives of tens of millions of Ukrainian people at home and abroad.

Among them are [the country's estimated 95,000 researchers](#). Until now, they were part of a modernizing scientific system that was beginning to throw off its Soviet-era shackles and integrate more closely with European research. Six months ago, there was a lot of interest in Ukraine and young people were heading up research departments, says George Gamota, a Ukrainian-born US physicist who left in 1944 and helped Ukraine to develop its scientific system after it gained independence in 1991. Now, the war has destroyed science centres in cities such as Kharkiv, Sumy and Mariupol and "a complete reconstruction will be needed once the war ends", says Gamota.

It is not yet possible to say how many researchers are casualties or have fled the war, although Gamota suggests that some 22,000 — mainly women with children — have left. Scientists worldwide have stepped up to help their colleagues through grassroots efforts such as #ScienceForUkraine, which has collated thousands of job offers at labs worldwide for Ukrainian researchers in need. Governments, universities and organizations such as CARA, the Council for At-Risk Academics in London, are also helping refugee scientists to resettle.

“There’s a lot of pressure in universities from academics and students” across the world, says Stephen Wordsworth, CARA’s executive director. “There’s a great awareness that there are people like them in other countries that are under considerable threat.” CARA is currently helping to place around 100 Ukrainian academics in research positions, mostly in the United Kingdom. “Many of them are optimistic in the circumstances. They’re thinking in terms of, ‘Maybe in six months’ time I’ll be able to go home again’,” he says.

And, in many cases, research and university teaching is continuing where possible at Ukrainian institutions, led by scientists who have remained at home or by refugee researchers who continue their work from overseas. “There’s quite a lot of work being done to keep universities functioning,” says Wordsworth.

Nature spoke to three Ukrainian researchers whose lives have been upended by Russia’s brutal invasion. Here are their stories.

‘I wanted to return power to my hands’

Olena Prysiazhna, plasma physicist, fled Ukraine for the Netherlands.

When the Russian invasion started, Prysiazhna didn’t think it would last. “We didn’t want to believe that this was happening. We always thought, ‘It’s going to be one or two weeks and this madness will be over’,” says Prysiazhna, who works at Taras Shevchenko National University of Kyiv.

Leaving their village, where her family had spent most of their lives, was difficult — emotionally and logistically. “We didn’t have a plan,” she says. “It’s an unusual feeling, when you don’t know where to go or what to do. Usually you have control of your life, but in war, you lose control of your life no matter what you do.”

Without access to a car, after about a week travelling inside the country, they found a driver who helped them get to the Polish border. They walked across and were helped by volunteers and taken to a refugee centre. From there, they took the next bus — regardless of destination — that would allow their

dog on board. (“He’s part of our family,” says Prysiazhna. “We couldn’t leave him behind.”)

They ended up in the Netherlands on 16 March. The first days were hard, says Prysiazhna. The trauma of war completely drained their energy, and she and her family initially managed only essential tasks, such as walking the dog. But after a while, Prysiazhna wanted to do something useful. “I wanted to return power to my hands.”



Plasma physicist Olena Prysiazhna (right) with her mother and sister.Credit: Ilvy Njiokiktjien/VII Photo for *Nature*

Prysiazhna contacted a Ukrainian scientist in the Netherlands, medical physicist Oleksandra Ivashchenko, who had studied at Prysiazhna’s university and had reached out to her to offer support. Ivashchenko, at Leiden University Medical Center, was helping to coordinate the #ScienceForUkraine effort.

With Ivashchenko's help, Prysiazhna and her sister, who is also a physicist, have begun visiting and talking to researchers at the Dutch Institute for Fundamental Energy Research in Eindhoven. It's still early days — Prysiazhna is exploring which direction her research could take and whether an arrangement might be formalized. One option might be to continue her work on optical emission spectroscopy analysis, a technique that can probe the properties of plasmas. "If we want to make everything better, we have to work. We have to make our small steps and do everything we can," says Prysiazhna.

She continues to teach online classes to her students in Ukraine. "Even during some research, I make a small break of one or two hours and conduct classes."

But there are challenges ahead — Prysiazhna is still trying to sort out her visa and work documents before she can settle in earnest, and the future is deeply uncertain. She follows news about Ukraine daily and hopes to return, but doesn't know whether she'll still have a home if she does.

Amid the trauma of the war and her flight, Prysiazhna has found some comfort in the people who have helped her. "I was surprised by how people can behave in a good way — how much good I got from people who I don't know," she says. "I can't even express how important it is."

But the brutality that this war has surfaced has shocked her deeply. "What's happening right now to some cities, I can't believe that someone can do that."

For now, Prysiazhna wants people to know about the horrors of this war. "I'd like to remind people, because silence is participation," she says. "The most important thing is being human. Don't ever forget."

'I will do anything to help Ukrainian scientists'

Taras Oleksyk, genomics researcher in Michigan, assisting people from Ukraine.



Taras Oleksyk was born in Ukraine and heads a genomics lab at Oakland University in Michigan. Credit: Emily Rose Bennett for *Nature*

This year, Taras Oleksyk is hoping to welcome to his laboratory a person with an unusual CV. Valerii Pokrytiuk, a young data engineer, has been accepted to do a master's in bioinformatics at Oleksyk's lab at Oakland University in Rochester, Michigan. But first, Pokrytiuk will serve his nation on the front line as a medic in the Ukrainian army.

Pokrytiuk won't be the only Ukrainian in the lab. Oleksyk was born in the country and left in 1992 to finish graduate school in the United States. A genomics researcher, he has since the early 2010s spearheaded an effort to chart the genome diversity of Russia and Ukraine, which he calls a "desert" in population genetics. (Tensions between the two nations have doomed that project, says Oleksyk, leading him to create a new one focusing on Ukrainian genomics.) But since Russia invaded his country of birth on 24 February, Oleksyk has devoted his work and personal life to a different purpose: helping Ukrainian researchers in any way he can.

When the war broke out, the first challenge was assisting one of his graduate students, Khrystyna Shchubelka, to get out of Ukraine, where she was organizing a genomics collaboration with a lab there. Shchubelka, who is Ukrainian, had also taken her baby to visit relatives. Her flights were cancelled and she had to walk across the border to Slovakia, says Oleksyk, who helped to raise money for flights to get her back to the United States.

He then began organizing assistance from Michigan with his wife, Audrey Majeske, who is also a biologist at Oakland. They set up an information table at the university where people could ask questions about the war, held a candlelight vigil, raised money for Ukrainian aid and hosted refugees at home. Oleksyk, who has a joint faculty appointment at Uzhhorod National University in western Ukraine, also teaches an online weekly university genomics course to 12 or so students in Ukraine, and the course material is shared with about 100 people. Sometimes the lectures are cancelled because of air raids, but “people log in from basements. They’re interested in genomics”, he says. “I am humbled by the defiance of Ukrainian students that keep coming to class in the time of war.”

Now, Oleksyk’s focus is on trying to find jobs or university places for Ukrainian researchers and students who have arrived in the United States. In addition to organizing a funded position in his lab for Pokrytiuk, he has secured a place for another Ukrainian refugee to begin graduate studies in biology, and is trying to convince the university to waive tuition fees for undergraduate students from the country. “I will do anything possible to provide opportunities to Ukrainian scientists,” he says.



Students from Ukraine and Puerto Rico with genomics researcher Taras Oleksyk (standing) in his lab. Credit: Emily Rose Bennett for *Nature*

In Oleksyk's lab, scientists from Ukraine will have a taste of the familiar. The lab has three Ukrainian graduate students, and nearby Detroit has one of the United States' biggest Ukrainian communities. Oleksyk, too, knows how it feels to be displaced. He established his lab at Oakland after he left Puerto Rico in 2018 following its devastation by Hurricane Maria. He brought his graduate students from the island. "My lab is referred to as 'The Ukaricans,'" says Oleksyk.

The work has been intense. "My wife says sometimes I spend a little too much time, but this is constantly on our minds." And his situation, he says, doesn't compare to that of his family in the Zakarpattia oblast of western Ukraine. He and his sister opened their house to displaced compatriots — each room now has a family living in it, some 25 people in all. And his sister, an endocrinologist, is employing as many people with medical

training as she can in her clinic there. “Right now, I’m dealing with survivor’s guilt. Whatever I’m doing, it’s never enough.”

Despite the horrors of the war, Oleksyk can see a sliver of hope for research. “Maybe this is a great opportunity to shake off that Soviet baggage and move on,” he says. When the war is over, Ukraine can finally integrate into the international community and “we can do the great things, including science”, says Oleksyk. “I know Ukraine is going to win this. I know Ukrainian people, they’re stubborn as hell.”

‘We must protect our ability to work as scientists’

Valerii Pokrytiuk, bionformatician, an enlisted medic in the Ukrainian army.

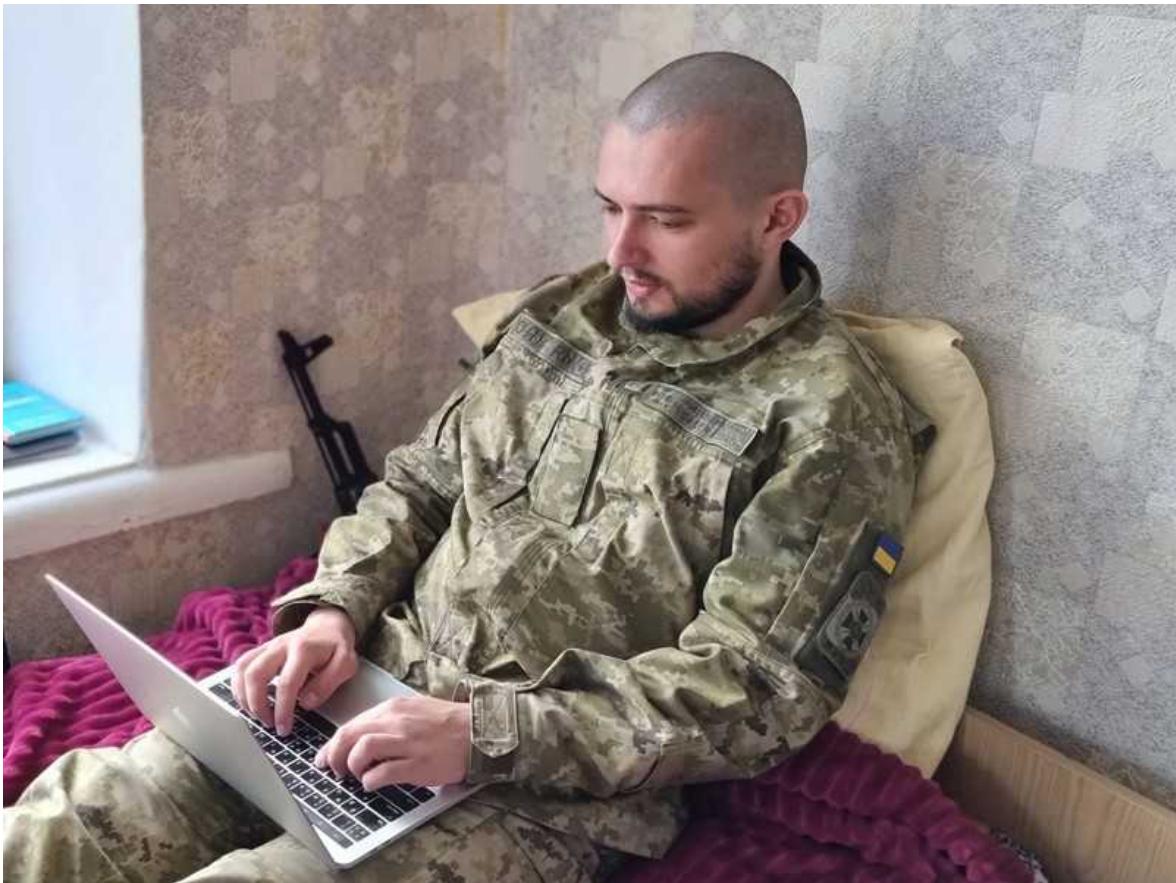


Valerii Pokrytiuk is a front-line medic who plans to do a master's in bioinformatics after the war.Credit: Valerii Pokrytiuk

In his small amount of free time, Valerii Pokrytiuk is doing a bioinformatics course online, and writing code on his smartphone. It's not as easy as doing it on a laptop, he says, but he's keen to learn computational biology.

Before the invasion, the 26-year-old was working as a data engineer for Emedgene, part of the US genetic-sequencing firm Illumina. When he spoke to *Nature* in early April, he was hours away from deploying to the front line as a medic in the Ukrainian army. Like many of his fellow citizens, he volunteered to fight for his nation on the first day of Russia's invasion. A few weeks later, Pokrytiuk, who has a degree in medicine, was called up and placed in a border-guard unit of the Ukrainian army. After three weeks of basic training, which included general tactics, shooting and tactical combat casualty care — a crash course in combat medicine — Pokrytiuk left for the front, although he could not reveal the location. “We are the first line of defence,” he says.

The decision to enlist wasn't difficult for Pokrytiuk, who was living and working in his university town of Vinnytsia in central Ukraine when the war started. “There are times to create science and there are other times to protect science and our ability to work as scientists,” says Pokrytiuk. “Today is the other kind of time,” he says. “I do like my science and coding routine more, but there was no choice for me, it was really simple.”



Valerii Pokrytiuk is doing an online course in bioinformatics in his small amount of free time.Credit: Valerii Pokrytiuk

Pokrytiuk knows that he won't be home in days or weeks. But he hopes it won't be years before the war ends. He has plans to go to the United States to get a master's degree in bioinformatics at Oakland University, working in Oleksyk's lab. Ultimately, he says, he wants to use computing and genomics research to save and prolong lives.

For now, Pokrytiuk shows little fear in the face of conflict. "I definitely don't want to lose my arms, because I'm a big fan of paragliding," he says. He has been more scared to think of what has happened in the besieged cities of Bucha, Irpin and Mariupol, which have seen intense fighting between Ukrainian and Russian forces and evidence of atrocities committed by Russian troops. Pokrytiuk's friends fought in some of those cities, and his relatives are in the military or have fled from occupied areas. "It's a personal story for me."

His immediate family, he says, is safe and his grandfather has refused to leave his home in northern Ukraine. “We joked a lot that my grandfather believes in the armed forces more than the Pope believes in God.”

Now Pokrytiuk is one of the people defending his grandfather and the rest of Ukraine. “I’m not scared,” he says. “We know what we’re fighting for.”

Nature **605**, 414-416 (2022)

doi: <https://doi.org/10.1038/d41586-022-01272-3>

Updates & Corrections

- **Correction 09 May 2022:** An earlier version of this story misidentified who opened up Oleksyk’s home to displaced people.

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-022-01272-3>

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

Books & Arts

- **Rise of drug-resistant tuberculosis is hidden in plain sight**

[16 May 2022]

Book Review • A history of TB infections homes in on India to illuminate the racism that denies treatment to millions around the world.

- **AI can burn or brighten, and intangible economics: Books in brief** [22 April 2022]

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

- BOOK REVIEW
- 16 May 2022

Rise of drug-resistant tuberculosis is hidden in plain sight

A history of TB infections homes in on India to illuminate the racism that denies treatment to millions around the world.

- [Heidi Ledford](#)



Tuberculosis thrives in Mumbai's cramped slums.Credit: Atul Loke/The New York Times/Redux/eyevine

The Phantom Plague: How Tuberculosis Shaped History *Vidya Krishnan* PublicAffairs (2022)

In 2020, while all eyes were on COVID-19, tuberculosis infected nearly 10 million people globally and killed 1.5 million. It was also the first year since 2005 that the number of deaths from the disease had risen. That increase was probably driven by COVID-19's impact on testing and treatment services.

The wealthy world still eagerly awaits a return to normal, pre-pandemic life. For many, “normal” was already deadly before 2020, global-health reporter Vidya Krishnan reminds us in *The Phantom Plague*. COVID-19 isn’t the only infectious respiratory disease knocking on the door: the threat of drug-resistant tuberculosis still hangs over the globe. As in the coronavirus pandemic, it is people with the least social, economic and historical capital who bear the brunt. “Poverty is the disease,” she writes; “TB the symptom.”

The greatest strength of *The Phantom Plague* is its highlighting of the forces that keep low- and middle-income countries hungry for medicines and at the mercy of *Mycobacterium tuberculosis*, the cause of tuberculosis. But first, the reader must meander through a familiar history of infectious-disease research, some only loosely tied to tuberculosis. Interesting nuggets do emerge: Dracula was an immigrant bearing a disease that polluted the blood of London residents; Arthur Conan Doyle might have been one of the first to recognize the spectre of drug resistance in tuberculosis, while researching a newspaper article.

The pay-off for persevering beyond those early chapters is worth the wait, however. Krishnan takes a chronological leap to recent years, and focuses on India, home to many of the world’s drug-resistant tuberculosis infections. She brings to life the darkness and stale air of life in Mumbai’s slums. There, seven-storey buildings are set just three metres apart — much closer than housing codes allow in areas away from these public-housing developments.

These buildings are hotbeds for tuberculosis. People who become infected, in Mumbai and across India, often wait months before they are properly diagnosed. In the meantime, they are given a hotchpotch of sometimes ineffective antibiotics, some of which have toxic side effects and nurture drug resistance.

This is an infuriating world, in which a tuberculosis infection that has spread to her ankle could threaten a young woman's life, and the old antibiotics available in India, such as kanamycin, do nothing to help and could ruin her hearing. Better antibiotics to treat drug-resistant tuberculosis in India are expensive and in limited supply. Until 2019, they were strictly rationed and available only to people who fitted a specific disease profile and lived near one of a handful of hospitals.

Krishnan rails against India's rationing of new tuberculosis drugs, such as bedaquiline, and backs up her arguments with horrifying personal stories. But here, the book's chronology can be confusing: for example, the rationing of bedaquiline is introduced and condemned, and Krishnan expresses bewildered outrage that the drug is, for a time, limited to those who live near certain hospitals. Only chapters later does she clearly lay out a key rationale for this restricted access: that researchers were still conducting trials to evaluate the drug's possible toxic effects on the heart. This organization of information creates some confusion.

Still, Krishnan makes a passionate case against the reasoning — trotted out all too often when it comes to treating infectious diseases in resource-poor regions — that people with tuberculosis cannot be trusted to take their medications and therefore should not be given the drugs they need. The conclusion of this flawed argument is that these newer, more effective drugs should be withheld from these populations because misuse could give rise to resistant pathogens that could then threaten richer countries. Krishnan argues effectively against this discrimination and labels it for what it is: racism.

She also takes on charities whose donations of crucial medicines, she says, foster dependence and allow countries to defer the need to establish sustainable supplies. She challenges patents and the biomedical monopolies they protect. Innovation scholars, predominantly at Western universities, spend careers analysing patent data and debating the relative values and costs of a strong patent system. Krishnan is not having it. She dubs support for strong international patents "fact free". To her, their only value is in wringing every cent from countries that lack the resources to fight back.

I sympathize with her passion. Her reporting has led her to people who have lost their hearing, their livelihoods, their loved ones — because, as she

argues, they were denied access to vital medicines produced in their own country. But I was disappointed to find no real rebuttal of the counterargument — that those medicines might not exist without the intellectual-property system that enables companies to profit from them. I yearned for her to take such arguments head-on and win.

The book is nonetheless a powerful look at the social determinants of health, and the lasting imprint of colonialism and segregation on public health. There is a desperate need for new drugs to combat drug-resistant tuberculosis. Meanwhile, as Krishnan reminds us, existing drugs are not being used effectively or fairly. It is this injustice that will feed the spread of drug-resistant tuberculosis.

Nature **605**, 417-418 (2022)

doi: <https://doi.org/10.1038/d41586-022-01342-6>

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-022-01342-6>

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

- BOOK REVIEW
- 22 April 2022

AI can burn or brighten, and intangible economics: Books in brief

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

- [Andrew Robinson](#) 0

SPARK

TIMOTHY J. JORGENSEN

Spark

Timothy J. Jorgensen *Princeton Univ. Press* (2022)

The use of electricity in medicine has long been controversial, notes health physicist Timothy Jorgensen. Eighteenth-century polymath Benjamin Franklin applied shocks to paralysed muscles with temporary success. In the 1930s, neurologist Ugo Cerletti pioneered painful but effective electroconvulsive therapy for schizophrenia and depression. Yet even today, “no one is sure exactly how ECT works”, says Jorgensen in his brilliant book. Now, business magnate Elon Musk plans to implant computer chips to treat brain disorders.

Ben Buchanan
and Andrew Imbrie

The New Fire

War, Peace, and
Democracy
in the Age of AI

The New Fire

Ben Buchanan & Andrew Imbrie *MIT Press* (2022)

Artificial intelligence (AI) is not like electricity, but like fire, say Ben Buchanan and Andrew Imbrie — academic specialists in emerging technology — in their authoritative, coruscating analysis of its current and future significance. Its potential impact ranges from illuminating to catastrophic, according to three rival and sometimes overlapping views from observers whom they label “evangelists, warriors and Cassandras”. “Three sparks ignite the new fire,” say the authors: data, algorithms and computing power.

Tomorrow's People

The Future of Humanity
in Ten Numbers

Paul Morland



1



43



121



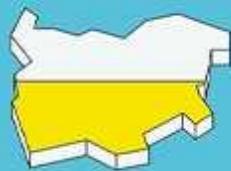
10



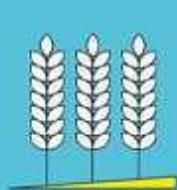
79,000



4,000,000,000



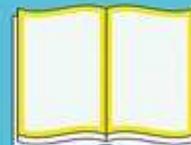
55



375



22



71

Tomorrow's People

Paul Morland *Picador* (2022)

“To most of us, the influence of demography on our future is far from obvious,” writes demographer Paul Morland. City dwellers tend to have low fertility, thereby creating an older population and eventually population decline, which could prompt migration and ethnic change, as in today’s United Kingdom — or might not, as in Japan. Morland’s careful book discusses ten indicators, one per chapter: infant mortality, population growth, urbanization, fertility, ageing, old age, population decline, ethnic change, education and food.

RESTARTING THE FUTURE



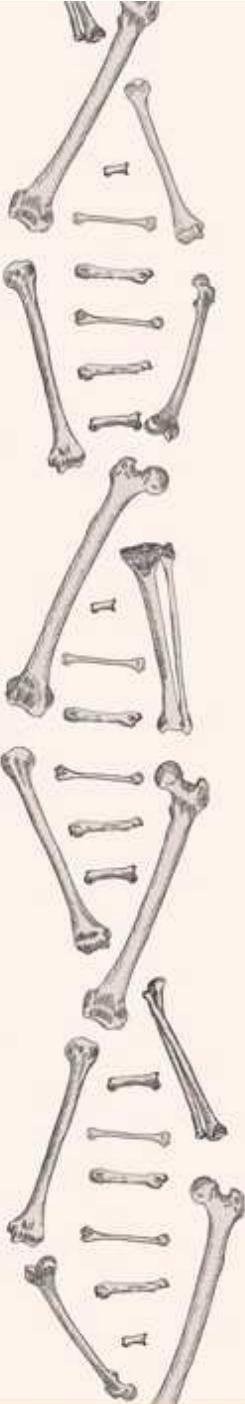
How to Fix the Intangible Economy

JONATHAN HASKEL
STIAN WESTLAKE

Restarting the Future

Jonathan Haskel & Stian Westlake *Princeton Univ. Press* (2022)

In the past few decades, growth has stagnated in advanced economies. This is odd, given low interest rates, high business profits and a wide belief that we live with “dizzying technological progress”, write economists Jonathan Haskel and Stian Westlake. They argue that the old economic model based on material production fails when it comes to intangible assets — such as software, data, design and business processes — that hinge on ideas, knowledge and relationships. Financial and state institutions must update to cope.



INEQUALITY

A Genetic History



CARLES LALUEZA-FOX

Inequality

Carles Lalueza-Fox *MIT Press* (2022)

Inequality and its origins will always preoccupy humans. In 2014, biologist Carles Lalueza-Fox led the retrieval of a genome from a European forager's skeleton more than 7,000 years old; his later studies revealed genetic evidence of "inequality and discrimination in different times and periods", as he describes in this significant book, written during the pandemic. He concludes by observing that COVID-19 has had an enhanced impact on poor people, which he anticipates will feature in future genetic studies.

Nature **605**, 418 (2022)

doi: <https://doi.org/10.1038/d41586-022-01130-2>

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-022-01130-2>

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

Opinion

- **[Want to prevent pandemics? Stop spillovers](#)** [12 May 2022]
Comment • Decision-makers discussing landmark agreements on health and biodiversity must include four actions to reduce the risk of animals and people exchanging viruses.
- **[One statistical analysis must not rule them all](#)** [17 May 2022]
Comment • Any single analysis hides an iceberg of uncertainty. Multi-team analysis can reveal it.

- COMMENT
- 12 May 2022

Want to prevent pandemics? Stop spillovers

Decision-makers discussing landmark agreements on health and biodiversity must include four actions to reduce the risk of animals and people exchanging viruses.

- [Neil M. Vora](#)⁰,
- [Lee Hannah](#)¹,
- [Susan Lieberman](#)²,
- [Mariana M. Vale](#)³,
- [Raina K. Plowright](#)⁴ &
- [Aaron S. Bernstein](#)⁵



Cattle encroach on the Amazon rainforest in Brazil as trees are burnt to clear land for grazing. Credit: Bruno Kelly/Reuters/Alamy

Spillover events, in which a pathogen that originates in animals jumps into people, have probably triggered every viral pandemic that's occurred since the start of the twentieth century¹. What's more, an August 2021 analysis of disease outbreaks over the past four centuries indicates that the yearly probability of pandemics could increase several-fold in the coming decades, largely because of human-induced environmental changes².

Fortunately, for around US\$20 billion per year, the likelihood of spillover could be greatly reduced³. This is the amount needed to halve global deforestation in hotspots for emerging infectious diseases; drastically curtail and regulate trade in wildlife; and greatly improve the ability to detect and control infectious diseases in farmed animals.

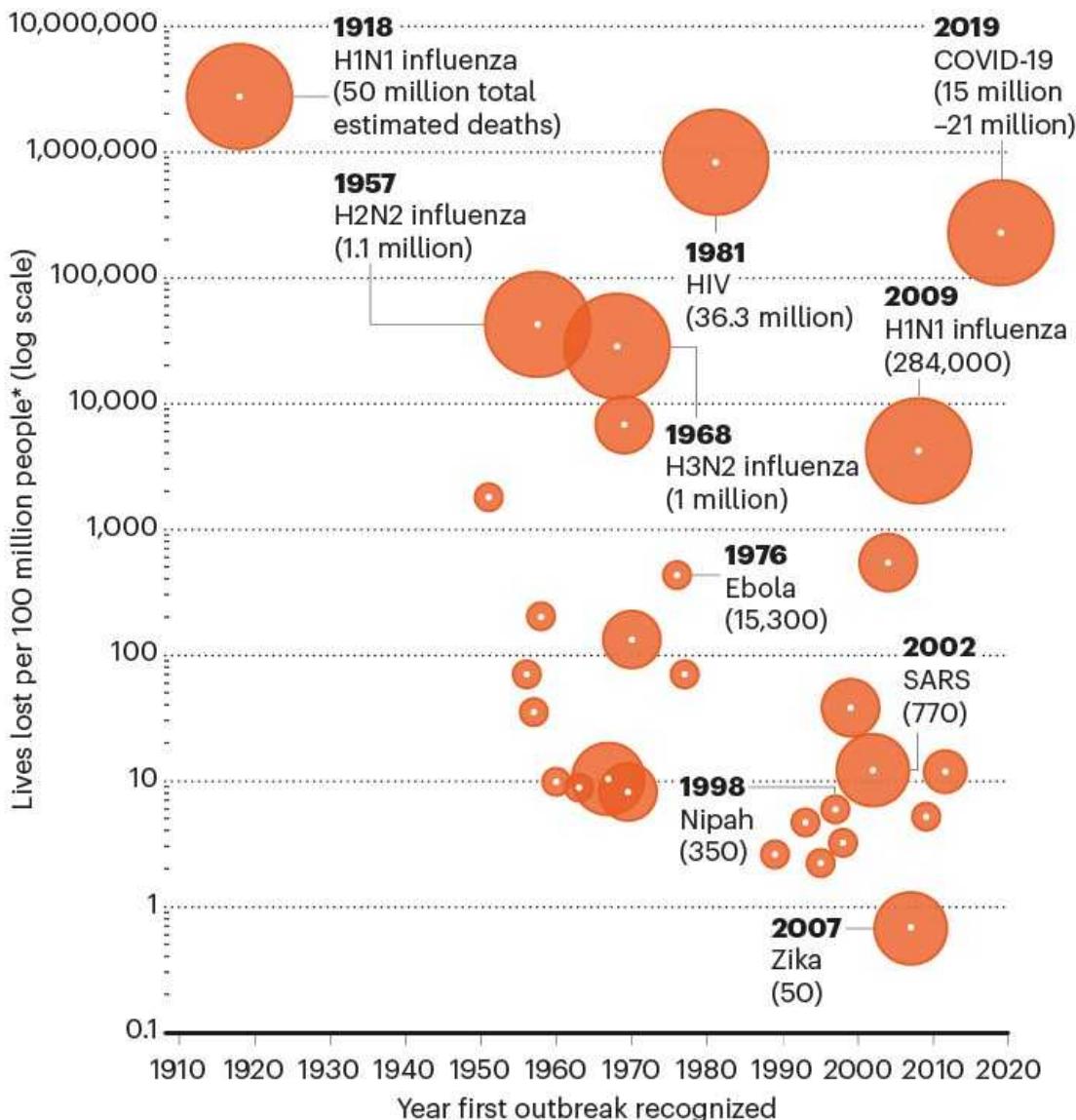
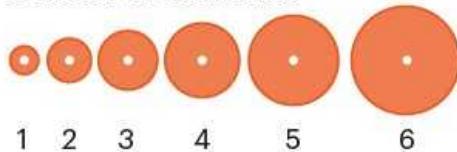
That is a small investment compared with the millions of lives lost and trillions of dollars spent in the COVID-19 pandemic. The cost is also one-twentieth of the statistical value of the lives lost each year to viral diseases

that have spilled over from animals since 1918 (see ‘Spillovers: a growing threat’), and less than one-tenth of the economic productivity erased per year¹.

SPILLOVERS: A GROWING THREAT

Deforestation and other changes have increased the likelihood of animal viruses jumping into people, with globalization and a higher density of human populations having increased the chance that such spillover events will be catastrophic. The annual economic loss from viral zoonoses since 1918 is US\$212 billion.

Number of continents



*Data are from viral outbreaks of zoonotic origin that resulted in ten or more deaths; figures in parentheses are total estimated deaths spread over multiple years (and over multiple outbreaks, in some instances) rounded to the nearest ten.

©nature

Source: Ref. 1

Yet many of the international efforts to better defend the world from future outbreaks, prompted by the COVID-19 pandemic, still fail to prioritize the prevention of spillover. Take, for example, the Independent Panel for Pandemic Preparedness and Response, established by the World Health Organization (WHO). The panel was convened in September 2020, in part to ensure that any future infectious-disease outbreak does not become another pandemic. In its 86-page report released last May, wildlife is mentioned twice; deforestation once.

We urge the decision-makers currently developing three landmark international endeavours to make the prevention of spillover central to each.

First, the G20 group of the world's 20 largest economies [provisionally agreed last month to create a global fund for pandemics](#). If realized, this could provide funding at levels that infectious-disease experts have been recommending for decades — around \$5 per person per year globally (see go.nature.com/3yjitwx). Second, an agreement to improve global approaches to pandemics is under discussion by the World Health Assembly (WHA), the decision-making body of the WHO. Third, a draft framework for biodiversity conservation — the post-2020 global biodiversity framework — is being negotiated by parties to the Convention on Biological Diversity.

Designed in the right way, these three international endeavours could foster a more proactive global approach to infectious diseases. This opportunity — to finally address the factors that drive major disease outbreaks, many of which also contribute to climate change and biodiversity loss — might not present itself again until the world faces another pandemic.

Four actions

The risk of spillover is greater when there are more opportunities for animals and humans to make contact, for instance in the trade of wildlife, in animal farming or when forests are cleared for mining, farming or roads. It is also more likely to happen under conditions that increase the likelihood of

infected animals shedding viruses – when they are housed in cramped conditions, say, or not fed properly.

Decades of research from epidemiology, ecology and genetics suggest that an effective global strategy to reduce the risk of spillover should focus on four actions^{1,3}.

First, tropical and subtropical forests must be protected. Various studies show that changes in the way land is used, particularly tropical and subtropical forests, might be the largest driver of emerging infectious diseases of zoonotic origin globally⁴. Wildlife that survives forest clearance or degradation tends to include species that can live alongside people, and that often host pathogens capable of infecting humans⁵. For example, in Bangladesh, bats that carry Nipah virus — which can kill 40–75% of people infected — now roost in areas of high human population density because their forest habitat has been almost entirely cleared⁶.

Furthermore, the loss of forests is driving climate change. This could in itself aid spillover by pushing animals, such as bats, out of regions that have become inhospitable and into areas where many people live⁷.

Yet forests can be protected even while agricultural productivity is increased — as long as there is enough political will and resources⁸. This was demonstrated by the 70% reduction in deforestation in the Amazon during 2004–12, largely through better monitoring, law enforcement and the provision of financial incentives to farmers. (Deforestation rates began increasing in 2013 due to changes in environmental legislation, and have [risen sharply since 2019 during Jair Bolsonaro's presidency](#).)

Second, commercial markets and trade of live wild animals that pose a public-health risk must be banned or strictly regulated, both domestically and internationally.

Doing this would be consistent with [the call made by the WHO and other organizations in 2021 for countries to temporarily suspend the trade in live caught wild mammals](#), and to close sections of markets selling such animals. Several countries have already acted along these lines. [In China, the trade and consumption of most terrestrial wildlife has been banned in response to](#)

[COVID-19](#). Similarly, Gabon has prohibited the sale of certain mammal species as food in markets.



A worker in a crowded chicken farm in Anhui province, China.Credit:
Jianan Yu/Reuters

Restrictions on urban and peri-urban commercial markets and trade must not infringe on the rights and needs of Indigenous peoples and local communities, who often rely on wildlife for food security, livelihoods and cultural practices. There are already different rules for hunting depending on the community in many countries, including Brazil, Canada and the United States.

Third, biosecurity must be improved when dealing with farmed animals. Among other measures, this could be achieved through better veterinary care, enhanced surveillance for animal disease, improvements to feeding and housing animals, and quarantines to limit pathogen spread.

Poor health among farmed animals increases their risk of becoming infected with pathogens — and of spreading them. And nearly 80% of livestock

pathogens can infect multiple host species, including wildlife and humans⁹.

Fourth, particularly in hotspots for the emergence of infectious diseases, people's health and economic security should be improved.

People in poor health — such as those who have malnutrition or uncontrolled HIV infection — can be more susceptible to zoonotic pathogens. And, particularly in immunosuppressed individuals such as these, pathogens can mutate before being passed on to others¹⁰.

What's more, some communities — especially those in rural areas — use natural resources to produce commodities or generate income in a way that brings them into contact with wildlife or wildlife by-products. In Bangladesh, for example, date palm sap, which is consumed as a drink in various forms, is often collected in pots attached to palm trees. These can become contaminated with bodily substances from bats. A 2016 investigation linked this practice to 14 Nipah virus infections in humans that caused 8 deaths¹¹.

Providing communities with both education and tools to reduce the risk of harm is crucial. Tools can be something as simple as pot covers to prevent contamination of date palm sap, in the case of the Bangladesh example.

In fact, providing educational opportunities alongside health-care services and training in alternative livelihood skills, such as organic agriculture, can help both people and the environment. For instance, the non-governmental organization Health in Harmony in Portland, Oregon, has invested in community-designed interventions in Indonesian Borneo. During 2007–17, these contributed to a 90% reduction in the number of households that were reliant on illegal logging for their main livelihood. This, in turn, reduced local rainforest loss by 70%. Infant mortality also fell by 67% in the programme's catchment area¹².

Systems-oriented interventions of this type need to be better understood, and the most effective ones scaled up.

Wise investment

Such strategies to prevent spillover would reduce our dependence on containment measures, such as human disease surveillance, contact tracing, lockdowns, vaccines and therapeutics. These interventions are crucial, but are often expensive and implemented too late — in short, they are insufficient when used alone to deal with emerging infectious diseases.

The COVID-19 pandemic has exposed the real-world limitations of these reactive measures — particularly in an age of disinformation and rising populism. For example, despite the US federal government spending more than \$3.7 trillion on its pandemic response as of the end of March, nearly one million people in the United States — or around one in 330 — have died from COVID-19 (see go.nature.com/39jtdfh and go.nature.com/38urqvc). Globally, between 15 million and 21 million lives are estimated to have been lost during the COVID-19 pandemic beyond what would be expected under non-pandemic conditions (known as excess deaths; see *Nature* <https://doi.org/htd6; 2022>). And a 2021 model indicates that, by 2025, \$157 billion will have been spent on COVID-19 vaccines alone (see go.nature.com/3jqds76).



A farmer in Myanmar gathers sap from a palm tree to make wine. Contamination of the collection pots with excretions from bats can spread diseases to humans. Credit: Wolfgang Kaehler/LightRocket via Getty

Preventing spillover also protects people, domesticated animals and wildlife in the places that can least afford harm — making it more equitable than containment. For example, almost 18 months since COVID-19 vaccines first became publicly available, only 21% of the total population of Africa has received at least one dose. In the United States and Canada, the figure is nearly 80% (see go.nature.com/3vrdpfo). Meanwhile, Pfizer's total drug sales rose from \$43 billion in 2020 to \$72 billion in 2021, largely because of the company's COVID-19 vaccine, the best-selling drug of 2021¹³.

Lastly, unlike containment measures, actions to prevent spillover also help to stop spillback, in which zoonotic pathogens move back from humans to animals and then jump again into people. Selection pressures can differ across species, making such jumps a potential source of new variants that can evade existing immunity. Some researchers have suggested that spillback was possibly responsible for the emergence of the Omicron variant of SARS-CoV-2 (see [Nature 602, 26–28; 2022](https://doi.org/10.1038/nature.2022.14710)).

Seize the day

Over the past year, the [administration of US President Joe Biden](#) and two international panels ([one established in 2020 by the WHO](#) and [the other in 2021 by the G20](#)) have released guidance on how to improve approaches to pandemics. All recommendations released so far acknowledge spillover as the predominant cause of emerging infectious diseases. None adequately discusses how that risk might be mitigated. Likewise, a PubMed search for the spike protein of SARS-CoV-2 yields thousands of papers, yet only a handful of studies investigate coronavirus dynamics in bats, from which SARS-CoV-2 is likely to have originated¹⁴.

Spillover prevention is probably being overlooked for several reasons. Upstream animal and environmental sources of pathogens might be being neglected by biomedical researchers and their funders because they are part of complex systems — research into which does not tend to lead to tangible,

profitable outputs. Also, most people working in public health and biomedical sciences have limited training in ecology, wildlife biology, conservation and anthropology.

There is growing recognition of the importance of cross-sectoral collaboration, including soaring advocacy for the '[One Health](#)' approach — [an integrated view of health that recognizes links between the environment, animals and humans](#). But, in general, this has yet to translate into action to prevent pandemics.

Another challenge is that it can take decades to realize the benefits of preventing spillover, instead of weeks or months for containment measures. Benefits can be harder to quantify for spillover prevention, no matter how much time passes, because, if measures are successful, no outbreak occurs. Prevention also runs counter to individual, societal and political tendencies to wait for a catastrophe before taking action.

The global pandemic fund, the WHA pandemic agreement and the post-2020 global biodiversity framework all present fresh chances to shift this mindset and put in place a coordinated global effort to reduce the risk of spillover alongside crucial pandemic preparedness efforts.

Global fund for pandemics

First and foremost, a global fund for pandemics will be key to ensuring that the wealth of evidence on spillover prevention is translated into action. Funding for spillover prevention should not be folded into existing conservation funds, nor draw on any other existing funding streams.

Investments must be targeted to those regions and practices where the risk of spillover is greatest, from southeast Asia and Central Africa to the Amazon Basin and beyond. Actions to prevent spillover in these areas, particularly by reducing deforestation, would also help to mitigate climate change and reduce loss of biodiversity. But conservation is itself drastically underfunded. As an example, natural solutions (such as conservation, restoration and improved management of forests, wetlands and grasslands) represent more than one-third of the climate mitigation needed by 2030 to

stabilize warming to well below 2 °C¹⁵. Yet these approaches receive less than 2% of global funds for climate mitigation¹⁶. (Energy systems receive more than half.)

In short, the decision-makers backing the global fund for pandemics must not assume that existing funds are dealing with the threat of spillover — they are not. The loss of primary tropical forest was 12% higher in 2020 than in 2019, despite the economic downturn triggered by COVID-19. This underscores [the continuing threat to forests](#).

Funding must be sustained for decades to ensure that efforts to reduce the risk of spillover are in place long enough to yield results.

WHA pandemic agreement

In 2020, the president of the European Council, Charles Michel, called for a treaty to enable a more coordinated global response to major epidemics and pandemics. Last year, more than 20 world leaders began echoing this call, and the WHA launched the negotiation of an agreement (potentially, a treaty or other international instrument) to “strengthen pandemic prevention, preparedness, and response” at the end of 2021.

Such a multilateral agreement could help to ensure more-equitable international action around the transfer of scientific knowledge, medical supplies, vaccines and therapeutics. It could also address some of the constraints currently imposed on the WHO, and define more clearly the conditions under which governments must notify others of a potential disease threat. The COVID-19 pandemic exposed the shortcomings of the International Health Regulations on many of these fronts¹⁷. (This legal framework defines countries’ rights and obligations in the handling of public-health events and emergencies that could cross borders.)

We urge negotiators to ensure that the four actions to prevent spillover outlined here are prioritized in the WHA pandemic agreement. For instance, it could require countries to create national action plans for pandemics that include reducing deforestation and closing or strictly regulating live wildlife markets. A reporting mechanism should also be developed to evaluate

progress in implementing the agreement. This could build on experience from existing schemes, such as the WHO Joint External Evaluation process (used to assess countries' capacities to handle public-health risks) and the verification regime of the Chemical Weapons Convention.

Commitments to expand pathogen surveillance at interfaces between humans, domesticated animals and wildlife — from US mink farms and Asian wet markets to areas of high deforestation in South America — should also be wrapped into the WHA agreement. Surveillance will not prevent spillover, but it could enable earlier detection and better control of zoonotic outbreaks, and provide a better understanding of the conditions that cause them. Disease surveillance would improve simply through investing in clinical care for both people and animals in emerging infectious-disease hotspots.

Convention on Biological Diversity

We are in the midst of the sixth mass extinction, and activities that drive the loss of biodiversity, such as deforestation, also contribute to the emergence of infectious disease. Meanwhile, epidemics and pandemics resulting from the exploitation of nature can lead to further conservation setbacks — because of economic damage from lost tourism and staff shortages affecting management of protected areas, among other factors¹⁸. Also, pathogens that infect people can be transmitted to other animals and decimate those populations. For instance, an Ebola outbreak in the Republic of Congo in 2002–03 is thought to have killed 5,000 gorillas¹⁹.

Yet the global biodiversity framework currently being negotiated by the Convention on Biological Diversity fails to explicitly address the negative feedback cycle between environmental degradation, wildlife exploitation and the emergence of pathogens. The first draft made no mention of pandemics. Text about spillover prevention was proposed in March, but it has yet to be agreed on.

Again, this omission stems largely from the siloing of disciplines and expertise. Just as the specialists relied on for the WHA pandemic agreement tend to be those in the health sector, those informing the Convention on

Biological Diversity tend to be specialists in environmental science and conservation.

The global biodiversity framework, scheduled to be agreed at the Conference of the Parties later this year, must strongly reflect the environment–health connection. This means explicitly including spillover prevention in any text relating to the exploitation of wildlife and nature’s contributions to people. Failing to connect these dots weakens the ability of the convention to achieve its own objectives around conservation and the sustainable use of resources.

Preventive health care

A reactive response to catastrophe need not be the norm. In many countries, preventive health care for chronic diseases is widely embraced because of its obvious health and economic benefits. For instance, dozens of colorectal cancer deaths are averted for every 1,000 people screened using colonoscopies or other methods²⁰. A preventive approach does not detract from the importance of treating diseases when they occur.

With all the stressors now being placed on the biosphere — and the negative implications this has for human health — leaders urgently need to apply this way of thinking to pandemics.

Nature **605**, 419–422 (2022)

doi: <https://doi.org/10.1038/d41586-022-01312-y>

References

1. Bernstein, A. S. *et al.* *Sci. Adv.* **8**, eabl4183 (2022).
2. Marani, M., Katul, G. G., Pan, W. K. & Parolari, A. J. *Proc. Natl Acad. Sci. USA* **118**, e2105482118 (2021).
3. Dobson, A. P. *et al.* *Science* **369**, 379–381 (2020).

4. Loh, E. H. *et al.* *Vector Borne Zoonotic Dis.* **15**, 432–437 (2015).
5. Gibb, R. *et al.* *Nature* **584**, 398–402 (2020).
6. McKee, C. D. *et al.* *Viruses* **13**, 169 (2021).
7. Carlson, C. J. *et al.* *Nature* <https://doi.org/10.1038/s41586-022-04788-w> (2022).
8. Nepstad, D. *et al.* *Science* **344**, 1118–1123 (2014).
9. Cleaveland, S., Laurenson, M. K. & Taylor, L. H. *Phil. Trans. R. Soc. B* **356**, 991–999 (2001).
10. Weigang, S. *et al.* *Nature Commun.* **12**, 6405 (2021).
11. Islam, M. S. *et al.* *Emerg. Infect. Dis.* **22**, 664–670 (2016).
12. Jones, I. J. *et al.* *Proc. Natl Acad. Sci. USA* **117**, 28515–28524 (2020).
13. Urquhart, L. *Nature Rev. Drug Discov.* **21**, 251 (2022).
14. Ruiz-Aravena, M. *et al.* *Nature Rev. Microbiol.* **20**, 299–314 (2022).
15. Griscom Bronson, W. *et al.* *Proc. Natl Acad. Sci. USA* **114**, 11645–11650 (2017).
16. Buchner, B. *et al.* *Global Landscape of Climate Finance 2021* (Climate Policy Initiative, 2021).
17. Singh, S. *et al.* *Lancet* **398**, 2109–2124 (2021).
18. Hockings, M. *et al.* *PARKS* **26**, 7–24 (2020).
19. Bermejo, M. *et al.* *Science* **314**, 1564–1564 (2006).
20. Knudsen, A. B. *et al.* *JAMA* **325**, 1998–2011 (2021).

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

- COMMENT
- 17 May 2022

One statistical analysis must not rule them all

Any single analysis hides an iceberg of uncertainty. Multi-team analysis can reveal it.

- [Eric-Jan Wagenmakers](#) ⁰,
- [Alexandra Sarafoglou](#) ¹ &
- [Balazs Aczel](#) ²

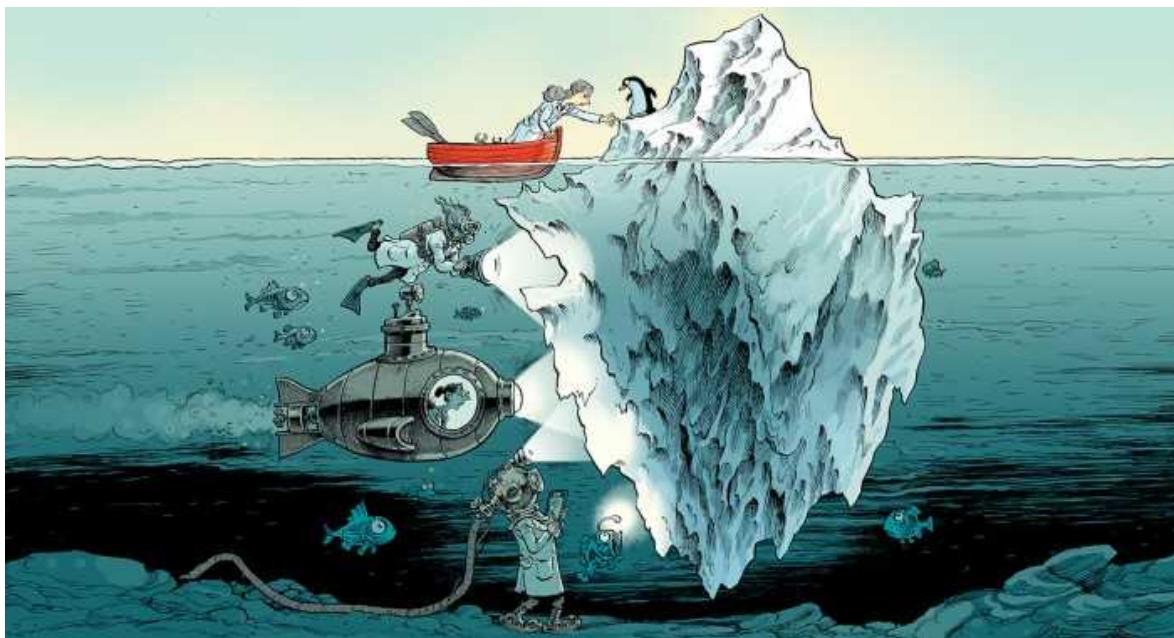


Illustration by David Parkins

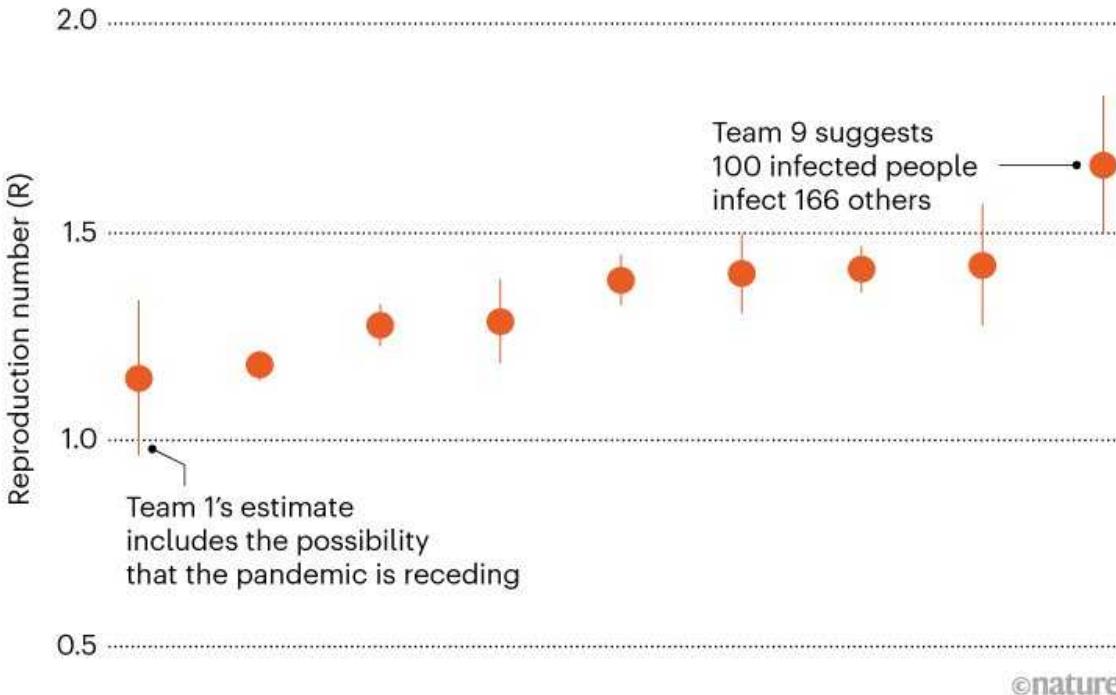
A typical journal article contains the results of only one analysis pipeline, by one set of analysts. Even in the best of circumstances, there is reason to think that judicious alternative analyses would yield different outcomes.

For example, in 2020, the UK Scientific Pandemic Influenza Group on Modelling asked nine teams to calculate the reproduction number R for COVID-19 infections¹. The teams chose from an abundance of data (deaths, hospital admissions, testing rates) and modelling approaches. Despite the clarity of the question, the variability of the estimates across teams was considerable (see ‘Nine teams, nine estimates’).

On 8 October 2020, the most optimistic estimate suggested that every 100 people with COVID-19 would infect 115 others, but perhaps as few as 96, the latter figure implying that the pandemic might actually be retreating. By contrast, the most pessimistic estimate had 100 people with COVID-19 infecting 166 others, with an upper bound of 182, indicating a rapid spread. Although the consensus was that the trajectory of disease spread was cause for concern, the uncertainty across the nine teams was considerably larger than the uncertainty within any one team. It informed future work as the pandemic continued.

NINE TEAMS, NINE ESTIMATES

Comparing models of the rate of COVID-19’s spread in the United Kingdom in early October 2020 revealed a degree of uncertainty masked by any one model.



Source: Ref. 1

Flattering conclusion

This and other ‘multi-analyst’ projects show that independent statisticians hardly ever use the same procedure^{2–6}. Yet, in fields from ecology to psychology and from medicine to materials science, a single analysis is considered sufficient evidence to publish a finding and make a strong claim.

Over the past ten years, the concept of *P*-hacking has made researchers aware of how the ability to use many valid statistical procedures can tempt scientists to select the one that leads to the most flattering conclusion. Less understood is how restricting analyses to a single technique effectively blinds researchers to an important aspect of uncertainty, making results seem more precise than they really are.

To a statistician, uncertainty refers to the range of values that might reasonably be taken by, say, the reproduction number of COVID-19 or the correlation between religiosity and well-being⁶, or between cerebral cortical thickness and cognitive ability⁷, or any number of statistical estimates. We argue that the current mode of scientific publication — which settles for a single analysis — entrenches ‘model myopia’, a limited consideration of statistical assumptions. That leads to overconfidence and poor predictions.

To gauge the robustness of their conclusions, researchers should subject the data to multiple analyses; ideally, these would be carried out by one or more independent teams. We understand that this is a big shift in how science is done, that appropriate infrastructure and incentives are not yet in place, and that many researchers will recoil at the idea as being burdensome and impractical. Nonetheless, we argue that the benefits of broader, more-diverse approaches to statistical inference could be so consequential that it is imperative to consider how they might be made routine.

Charting uncertainty

Some 100 years ago, scholars such as Ronald Fisher advanced formal methods for hypothesis testing that are now considered indispensable for drawing conclusions from numerical data. (The *P* value, often used to determine ‘statistical significance’, is the best known.) Since then, a plethora

of tests and methods have been developed to quantify inferential uncertainty. But any single analysis draws on a very limited range of these. We posit that, as currently applied, uncertainty analyses reveal only the tip of the iceberg.

The dozen or so formal multi-analyst projects completed so far (see Supplementary information) show that levels of uncertainty are much higher than that suggested by any single team. In the 2020 Neuroimaging Analysis Replication and Prediction Study², 70 teams used the same functional magnetic resonance imaging (MRI) data to test 9 hypotheses about brain activity in a risky-decision task. For example, one hypothesis probed how a brain region is activated when people consider the prospect of a large gain. On average across the hypotheses, about 20% of the analyses constituted a ‘minority report’ with a qualitative conclusion opposite to that of the majority. For the three hypotheses that yielded the most ambiguous outcomes, around one-third of teams reported a statistically significant result, and therefore publishing work from any of one these teams would have hidden considerable uncertainty and the spread of possible conclusions. The study’s coordinators now advocate that multiple analyses of the same data be done routinely.

Another multi-analyst project was in finance³ and involved 164 teams that tested 6 hypotheses, such as whether market efficiency changes over time. Here again, the coordinators concluded that differences in findings were due not to errors, but to the wide range of alternative plausible analysis decisions and statistical models.

All of these projects have dispelled two myths about applied statistics. The first myth is that, for any data set, there exists a single, uniquely appropriate analysis procedure. In reality, even when there are scores of teams and the data are relatively simple, analysts almost never follow the same analytic procedure.

The second myth is that multiple plausible analyses would reliably yield similar conclusions. We argue that whenever researchers report a single result from a single statistical analysis, a vast amount of uncertainty is hidden from view. And although we endorse recent science-reform efforts, such as large-scale replication studies, preregistration and registered reports, these initiatives are not designed to reveal statistical fragility by exploring

the degree to which plausible alternative analyses can alter conclusions. In summary, formal methods, old and new, cannot cure model myopia, because they are firmly rooted in the single-analysis framework.

We need something else. The obvious treatment for model myopia is to apply more than one statistical model to the data. High-energy physics and astronomy have a strong tradition of teams carrying out their own analyses of other teams' research once the data are made public. Climate modellers routinely perform 'sensitivity analyses' by systematically removing and including variables to see how robust their conclusions are.

For other fields to make such a shift, journals, reviewers and researchers will have to change how they approach statistical inference. Instead of identifying and reporting the result of a single 'correct' analysis, statistical inference should be seen as a complex interplay of different plausible procedures and processing pipelines⁸. Journals could encourage this practice in at least two ways. First, they could adjust their submission guidelines to recommend the inclusion of multiple analyses (possibly reported in an online supplement)⁹. This would motivate researchers to either conduct extra analyses themselves or to recruit more analysts as co-authors. Second, journals could invite teams to contribute their own analyses in the form of comments on a recently accepted article.

False alarm?

Certainly, large-scale changes in how science is done are possible: expectations surrounding the sharing of data are growing. Medical journals now require that clinical trials be registered at launch for the results to be published. But proposals for change inevitably prompt critical reactions. Here are five that we've encountered.

Won't readers get confused? Currently, there are no comprehensive standards for, or conventions on, how to present and interpret the results of multiple analyses, and this situation could complicate how results are reported and make conclusions more ambiguous. But we argue that potential ambiguity is a key feature of multi-team analysis, not a bug. When conclusions are supported only by a subset of plausible models and analyses,

readers should be made aware. Facing uncertainty is always better than sweeping it under the rug.

Aren't other problems more pressing? Problems in empirical science include selective reporting, a lack of transparency around analyses, hypotheses that are divorced from the theories they are meant to support, and poor data sharing. It is important to make improvements in these areas — indeed, how data are collected and processed, and how variables are defined, will greatly influence all subsequent analyses. But multi-analyst approaches can still bring insight. In fact, multi-analyst projects usually excel in data sharing, transparent reporting and theory-driven research. We view the solutions to these problems as mutually reinforcing rather than as a zero-sum game.

Is it really worth the time and effort? Even those who see benefit in multiple analyses might not see a need for them to happen at the time of publication. Instead, they would argue that the original team be encouraged to pursue multiple analyses or that shared data can be reanalysed by other interested researchers after publication. We agree that both would be an improvement over the status quo (sensitivity analysis is a severely underused practice). However, they will not yield the same benefits as multi-team analyses done at the time of publication.

Post-publication analyses are usually published only if they drastically undercut the original conclusion. They can give rise to squabbles more than constructive discussion, and would come out after the authors and readers have already drawn conclusions based on a single analysis. Information about uncertainty is most useful at the time of analysis. However, we doubt whether a single team can muster the mental fortitude needed to reveal the fragility of their findings; there might be a strong temptation to select those analyses that, together, present a coherent story. In addition, a single research team usually has a somewhat narrow expertise in data analysis. For instance, each of the nine teams that produced different estimates for R would probably feel uncomfortable if they had to code and produce estimates using the other teams' models. Even for simple statistical scenarios (that is, a comparison of two outcomes — such as the proportions of people who improve after receiving a drug or placebo — and a test of a

linear correlation), several teams can apply widely divergent statistical models and procedures¹⁰.

Some sceptics doubt that multi-team analyses will consistently find broad enough ranges of results to make the effort worthwhile. We think that the outcomes of existing multi-analyst projects counter that argument, but it would be useful to gather evidence from yet more projects. The more multi-analyst approaches are undertaken, the clearer it will be as to how and when they are valuable.

Won't journals baulk? One sceptical response to our proposal is that multi-analyst projects will take longer, be more complicated to present and assess, and will even require new article formats — complications that will make journals reluctant to embrace the idea. We counter that the review and publication of a multi-analyst paper do not require a fundamentally different process. Multi-team projects have been published in a variety of journals, and most journals already publish comments attached to accepted manuscripts. We challenge journal editors to give multi-analyst projects a chance. For instance, editors might test the waters by organizing a special issue consisting of case studies. This should make it readily apparent whether the added value of the multi-analyst approach is worth the extra effort.

Won't it be a struggle to find analysts? One response to our proposal is that the bulk of multi-team analyses published so far are the product of demonstration projects wrapped into a single paper. These papers encompass several analyses with long author lists comprised mainly of enthusiasts for reform; most other researchers would see little benefit in being a minor contributor to a multi-analyst paper, especially one at the periphery of their core research interest. But we think enthusiasm has a broad base. In our multi-analyst projects, we have been known to receive more than 700 sign-ups in about 2 weeks.

Moreover, a range of incentives could attract teams of analysts, such as gaining co-authorship and the chance to work on important questions or simply to collaborate with specialists. Further incentives and catalysts are easy to imagine. In a forthcoming special issue of the journal *Religion, Brain & Behavior*, several teams will each publish their own conclusions and

interpretations of the research question addressed by the main article⁶, and this means each teams' contribution is individually recognized. When a question is particularly urgent, journals, governments and philanthropists should actively recruit or support multi-analysis teams.

Yet another approach would be to incorporate multiple analyses into training programs, which would be both useful for the research community and eye-opening for statisticians. (At least one university has incorporated replication studies into its curricula¹¹.) Ideally, participating in multiple analyses will be seen as part of being a good science 'citizen', and be rewarded through better prospects for hiring and promotion.

Whatever the mix of incentives and formats, the more that multiple analyses efforts are implemented and discussed, the easier they will become. What makes such multi-team efforts work well should be studied and applied to improve and expand the practice. As the scientific community learns how to run multi-team analyses and what can be learnt, acceptance and enthusiasm will grow.

We argue that rejecting the multi-analyst vision would be like Neo opting for the blue pill in the film *The Matrix*, and so continuing to dream of a reality that is comforting but false. Scientists and society will be better served by confronting the potential fragility of reported statistical outcomes. It is crucial for researchers and society to have an indication of such fragility from the moment the results are published, especially when these results have real-world ramifications. Recent many-analyst projects suggest that any single analysis will yield conclusions that are overconfident and unrepresentative. Overall, the benefit of increased insight will outweigh the extra effort.

Nature **605**, 423-425 (2022)

doi: <https://doi.org/10.1038/d41586-022-01332-8>

References

1. Scientific Pandemic Influenza Group on Modelling. *SPI-M-O: Consensus statement on COVID-19*, 8 October 2020 (2020).
2. Botvinik-Nezer, R. *et al.* *Nature* **582**, 84–88 (2020).
3. Menkveld, A. J. *et al.* Preprint at SSRN
<https://doi.org/10.2139/ssrn.3961574> (2021).
4. Silberzahn, R. *et al.* *Adv. Methods Pract. Psychol. Sci.* **1**, 337–356 (2018).
5. Steegen, S., Tuerlinckx, F., Gelman, A. & Vanpaemel, W. *Perspect. Psychol. Sci.* **11**, 702–712 (2016).
6. Hoogeveen, S. *et al.* Preprint at PsyArXiv
<https://doi.org/10.31234/osf.io/pbfye> (2022).
7. Marek, S. *et al.* *Nature* **603**, 654–660 (2022).
8. Wagenmakers, E.-J. *et al.* *Nature Hum. Behav.* **5**, 1473–1480 (2021).
9. Aczel, B. *et al.* *eLife* **10**, e72185 (2021).
10. van Dongen, N. N. N. *et al.* *Am. Stat.* **73**, 328–339 (2019).
11. Button, K. *Nature* **561**, 287 (2018).

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-022-01332-8>

Work

- **Childcare crowdfunding campaigns aim to keep mums on the academic track** [09 May 2022]

Career Feature • The lack of affordable childcare threatens the retention of scientist parents, especially mothers and those of colour.

- **Immune to frustrations** [10 May 2022]

Where I Work • Mercedes Segovia researches the role of dendritic cells in organizing the immune system.

- CAREER FEATURE
- 09 May 2022
- Correction [12 May 2022](#)

Childcare crowdfunding campaigns aim to keep mums on the academic track

The lack of affordable childcare threatens the retention of scientist parents, especially mothers and those of colour.

- [Virginia Gewin](#) ⁰

[Find a new job](#)



For financial support, researcher parents are turning to crowdfunding programmes — some of which target mothers of colour specifically. Credit: Getty

As Ava, a postdoctoral researcher at a top-tier university in North Carolina, bounces her three-month-old daughter on her lap during the final days of her family leave, she worries about whether she will be able to realize her dream of becoming a principal investigator (PI). Her biggest roadblock is the lack of affordable childcare. “You have to be competitive in order to get the PI positions, and I won’t be able to do that if I’m not able to put her in day care,” she says.

Ava’s only realistic option will cost roughly US\$1,700 a month — 49% of her monthly salary — for day care and fuel for her two- to three-hour commute to and from their home in an affordable neighborhood. Using a large slice of household income for day care — while missing out on many childhood milestones — doesn’t make financial or emotional sense. It’s a dilemma that many working mothers face: why bother working if they are just working to pay for childcare? To finish her postdoc, Ava worked out the following schedule with her PI: she will arrive at work by 5 a.m. to generate data, then leave at around 2 p.m. to spend quality time with her daughter, and she will analyse the data during evenings and weekends. “My whole goal is to be the best mom I can be, and if that means I have to put my career dream on the backburner, I may have to bite that bullet,” she says.



Fernanda Staniscuaski, founder of Brazil's Parent in Science movement, is crowdfunding for a second year to support early-career mothers.Credit: Flavio Dutra/Federal University of Rio Grande do Sul

Spurred by similar stories from scientist parents in the United States, where there is no national policy for paid family leave, Frankie Heyward, a postdoc studying neuro-epigenetics at Beth Israel Deaconess Medical Center in Boston, Massachusetts, wanted to help. As founder and current president of the National Black Postdoctoral Association (NBPA), Heyward started a crowdfunding initiative in February to help offset childcare costs for Black early-career scholars such as Ava. These scientist parents are so close to finishing their training, says Heyward. "They just need resources for childcare." The NBPA Childcare Fund has so far raised more than US\$5,500.In the past few years, several childcare crowdfunding campaigns have emerged in the United States and Brazil. Some of these initiatives are dedicated solely to mothers of colour, and others prioritize them. *Nature* highlights four such efforts here, but is keen to hear about others.

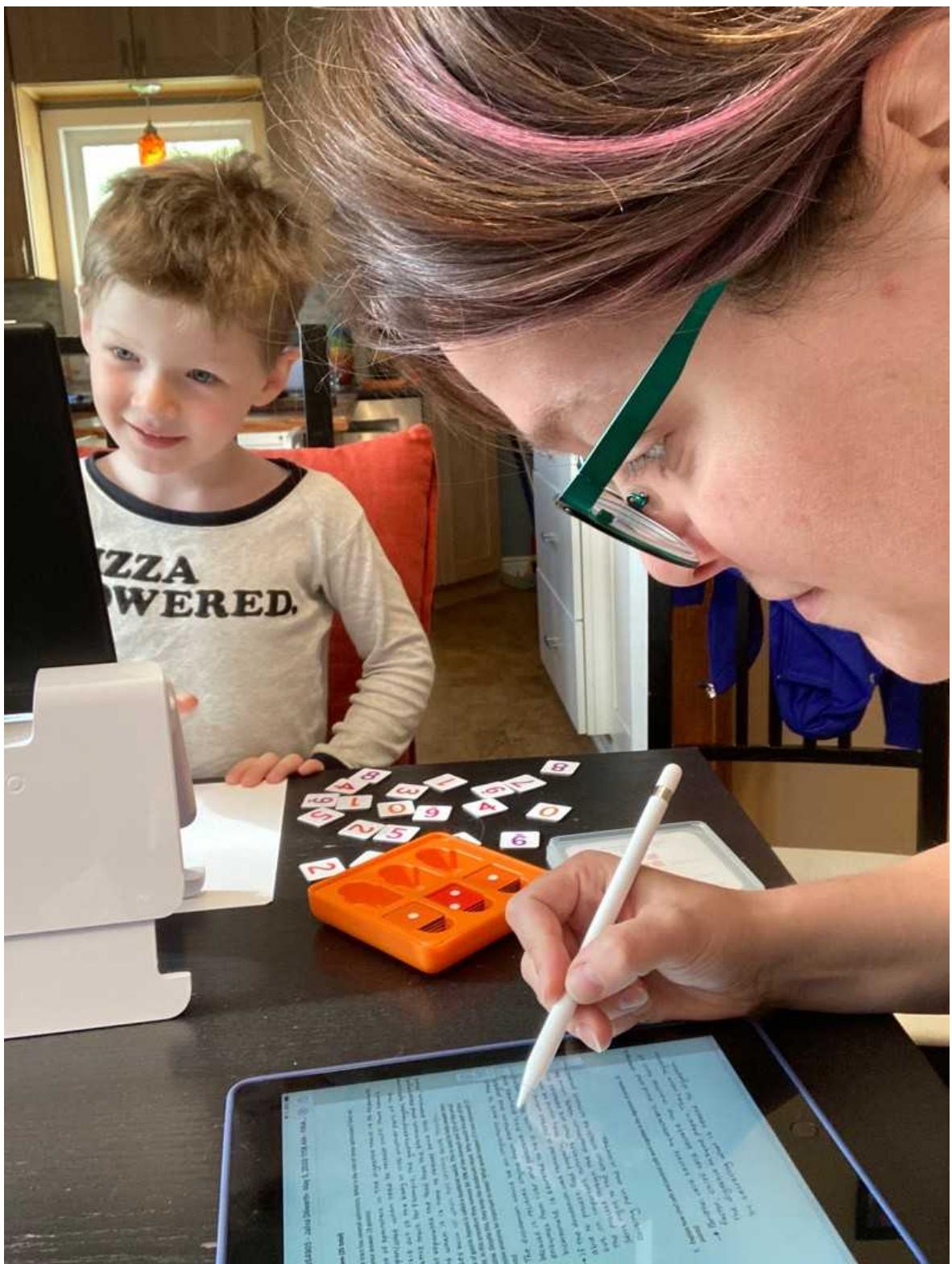
These grass-roots initiatives aim to help retain mothers in academia — loss of whom threatens to erase slim diversity gains in academic science. Ava,

who is 33 and requested that we don't use her real name to protect her privacy, is the average age at which male and female academics in the United States and Canada become parents¹. However, many mothers' careers are derailed. "We are losing women, and it is very frustrating to watch really good colleagues leave academia," says Mariya Sweetwyne, a renal-cell biologist at the University of Washington School of Medicine in Seattle and a founder of a crowdfunding campaign. "I'm so grateful of all the people who donated, but it's outrageous that we need to solicit private donations to offer a lifeline to a minoritized mother's career," she says. "If institutions are serious about their stated commitment to equity and diversifying academic medicine and research, then this lack of parenting support has to be addressed immediately."

Only about half of public US universities that grant bachelor's degrees offer on-site childcare — and even fewer private institutions do, according to 2021 analyses² conducted by Child Trends, a non-profit children's research organization in Bethesda, Maryland. Where they exist, university day-care centres are often expensive and have waiting lists of months to years. Yet, women report that paid parental leave and adequate childcare are crucial factors in applying for and staying in jobs.

Pandemic crowdfunding

The COVID-19 pandemic amplified existing barriers to women's careers. Many mothers assumed greater childcare responsibilities — and support became less reliable³. The lack of affordable childcare is an unrelenting concern amid efforts to improve diversity, equity and inclusivity on university campuses. "These crowdfunds highlight that women, particularly women of colour, are not getting the support that they need to survive, let alone thrive," in science, technology, engineering and maths, says Michelle Cardel, a health-outcomes researcher at the University of Florida in Gainesville, who studies gender disparities in academia.



Cynthia Harley works from home with her son. She created the Academic Mamas Foundation to support mums.Credit: Cynthia Harley

Last year, Sweetwyne organized an effort to help researcher mothers from under-represented groups through Academic Research Moms (ARM), a private Facebook group of roughly 3,100 women. Through the GoFundMe platform, ARM has so far raised more than \$90,000 specifically to offset childcare costs for medical and research trainees of colour. Following the 2020 death of George Floyd, a Black man who was murdered by a white police officer in Minneapolis, Minnesota, the group's conversations focused on the sluggish pace of increasing diversity in higher education. Sweetwyne thought a diversity-focused fund could at least serve to patch one hole in the leaky pipeline. Donors agreed — they raised \$60,000 in the first two months.

In many instances, the donations come from other academic mums, a group hit hard by pandemic strain. The irony is not lost on Sweetwyne that individual mothers are trying to help academic institutions fix their pipeline issues. “We’re not doing it for [the institutions]. We’re doing it for us,” she says of the almost 1,000 donors to the fund. Institutions, she adds, should do much more to address ongoing childcare concerns.

After exploring options to dispense donations, Sweetwyne found an enthusiastic partner in Meharry Medical College, a historically Black medical school in Nashville, Tennessee, with a strong need. There is no childcare on campus, and the average debt of students at Meharry is more than \$300,000, according to director of estate planning and giving Gloria Sanders. “Childcare continues to be a major need for employees and students,” she says. ARM funding has provided one-time awards of roughly \$1,250–1,500 to more than 30 women since 2020.

During the final year of her cancer-biology PhD studies at Meharry, Heather Beasley received an ARM award of \$1,500 — just shy of her monthly salary. “It was life-changing for me,” she says, explaining that rent plus the cost of her daughter’s childcare ate up most of her monthly income. “For one month, it offset major costs that I didn’t have to put on a credit card.”

Beasley estimates that she kept her daughter in childcare for 40% of her time in graduate school, in part because it was too expensive. The rest of the time, she worked nights and weekends, when her husband could look after their child. Once she finishes her current postdoctoral position at Vanderbilt

University in Nashville, [she intends to pursue higher-paying employment in industry](#). “I can influence my community without being subjected to low wages,” she says.



President of the US National Black Postdoctoral Association Frankie Heyward started a crowdfund to help retain mothers of colour in academia. Credit: Frankie Heyward

Female scientists are among the most difficult groups to retain in academia. Black and Hispanic women comprise just 5% or less of tenured faculty in the United States, according to a study⁴ that Cardel published in 2020. And the pandemic exacerbated the challenges. “When childcare was unavailable or limited, we saw a secondary epidemic of women leaving the workforce,” says Cardel. When it comes to institutional diversity, equity and inclusivity aims, Sweetwyne says, “it seems the goal is to increase the flow; it doesn’t seem like the goal is to reduce the leak”.

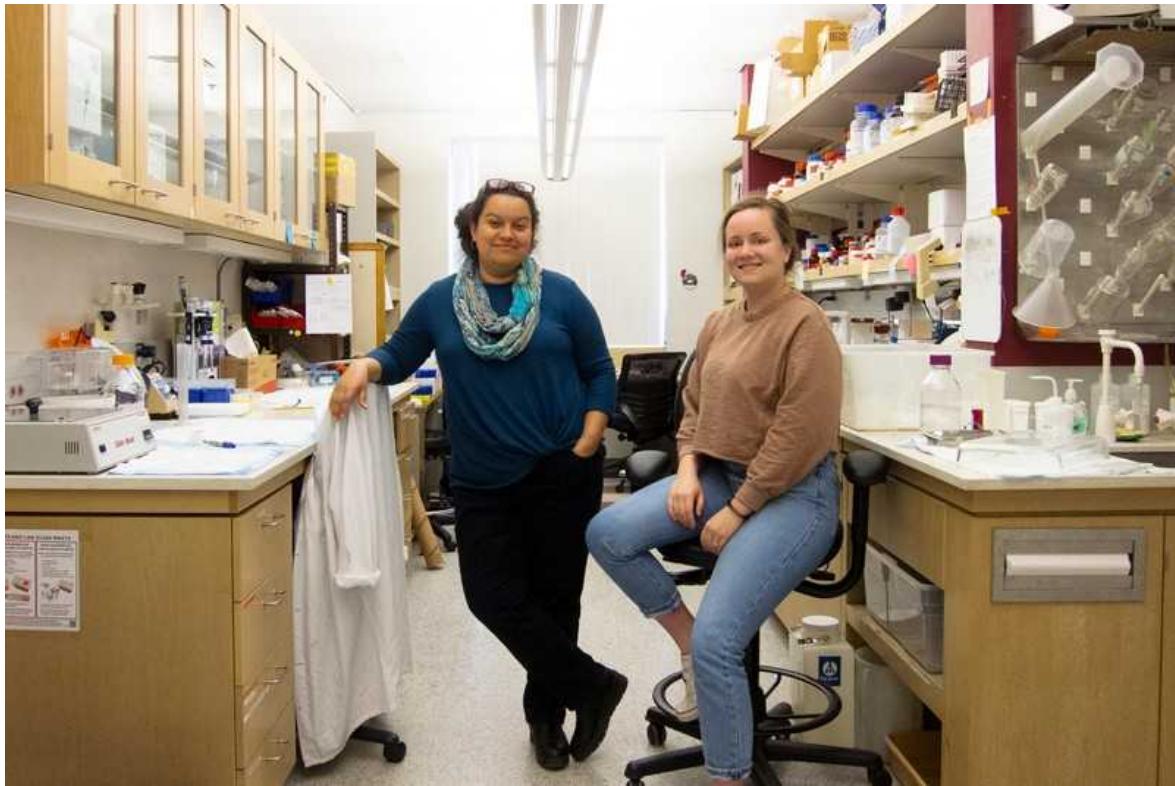
To try to curb that leak, Cindy Harley, a biologist at Metropolitan State University in Saint Paul, Minnesota, founded the Academic Mamas Foundation in 2018, and serves as the president of the non-profit

organization. The foundation has raised roughly \$12,000 to help support mothers, and has formed a partnership with the International Association of Maternal Action and Scholarship, an advocacy group. They distribute mainly informal small grants — \$250–500 at a time — to help people who have, for example, taken leave to care for children. Larger grants have funded childcare for conference attendees. But this ‘moms helping moms’ approach has been stymied by pandemic fatigue, says Harley.

Brazil’s crowdfunding success

In Brazil, the Parent in Science Movement gathered donations early in the pandemic for a similar crowdfunded effort, called the Tomorrow Program. The effort raised 100,000 Brazilian reais (US\$20,000) to offer doctoral-student mothers from minority ethnic groups [childcare stipends of roughly 700 reais \(US\\$150\) a month](#). They had 750 applicants for 26 slots.

Last year, the movement won a US\$40,000 Nature Research Award for Inspiring Women in Science. It plans to use around 20% of the prize money — plus crowdfunding that it hopes will triple that amount — to fund the Tomorrow Program for a second year. “While we realize crowdfunding is not a sustainable solution, it is the only alternative” because no private sponsors stepped up, says movement founder Fernanda Staniscuaski, a molecular biologist at the Federal University of Rio Grande do Sol in Porto Alegre, Brazil. This year, the fund will focus on undergraduate mothers and is accepting applications [during the last two weeks of May](#).



Mariya Sweetwyne (left), with laboratory technician Carolyn Mann, has raised more than \$90,000 for medical and research trainee mums of colour. Credit: Mariya Sweetwyne

Despite [significant cuts to Brazil's federal scientific agencies in recent years](#), some universities have created structural support. The State University of Rio de Janeiro, Brazil, pays 900 Brazilian reais monthly to student parents. “There is a big demand for this kind of action. It would be really nice if we had federal government support,” says Staniscuaski.

Although crowdfunding childcare is a stop-gap attempt to staunch the flow of scientist mothers, especially those from under-represented groups, from leaving academia, long-term measures are needed. A 2022 perspective piece⁵ published in Academic Medicine called for childcare stipends as part of US National Institutes of Health grants to address persistent racial and gender disparities in physician-scientists. “Grants should provide childcare stipends to PIs and trainees alike to ease the financial burden and show moms that they are valued in academia,” says study co-author Heather Ward, a psychiatrist and research fellow at Beth Israel Deaconess Medical Center.

One tenured scientist mother turned to grants to create lasting change. Lesley Lutes, director of clinical training for the doctoral programme in clinical psychology at the University of British Columbia's Okanagan Campus in Kelowna, Canada, channelled her own frustrations, including a three-year waiting list for childcare, into action. She partnered with the university to write two successful grant applications, totalling \$2.2 million, to triple the size of the existing non-profit childcare centre on campus to around 100 spots. Before the expansion, some women took extended leave, considered leaving academia or found more affordable childcare in nearby cities, she explains. "There wasn't more action before because women weren't at the table," says Lutes.

Some institutions are making affordable childcare a cornerstone of their recruitment and retention efforts. In 2019, a taskforce at Vanderbilt University outlined how to expand childcare access on campus. Their on-campus childcare facility currently costs around 10% below the market rate in Nashville and offers discounts on the basis of income level. Still, children of graduate and professional students hold only 18% of day-care slots.

"Our biggest hurdle has been recruiting and keeping childcare staff," says Patrick Retton, associate vice-chancellor for administration and chief of operations at Vanderbilt. To bolster that staff, the facility director secured two grants to provide salary bonuses. "Vanderbilt is focused on employee recruitment and retention, and one of the subtopics is the rewards and benefits, including childcare, to the university as a whole," says Retton.



Lesley Lutes (left), who obtained funding to expand the day-care centre at the University of British Columbia Okanagan Campus, with day-care manager Kim Letiec. Credit: Nathan Skolski/UBC

Vanderbilt's efforts to provide affordable, accessible childcare to retain academic parents is an obvious yet rare example. "[Academic mothers] are getting to the final phase of training and having to abandon their dreams because of the financial hurdles," says Heyward. "I want to show with my fund that when you provide support to individuals, they can stay the course," he adds, "and I want institutions to appreciate that."

Sweetwyne agrees: "We can't GoFundMe forever, but we do intend to keep hammering at this until — institutions willing — it's unnecessary."

Nature **605**, 575-577 (2022)

doi: <https://doi.org/10.1038/d41586-022-01289-8>

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

encouraged.

Updates & Corrections

- **Correction 12 May 2022:** An earlier version of this feature did not give the most up-to-date figures about Vanderbilt University's day-care facility, and erroneously stated that the two grants it received to provide salary bonuses were from the federal government.

References

1. Morgan, A. C. *et al. Sci. Adv.* **7**, eabd1996 (2021).
2. Ryberg, R., Rosenberg, R. & Warren, J. *Higher Education Can Support Parenting Students and Their Children with Accessible, Equitable Services* (Child Trends, 2021).
3. Yavorsky, J. E., Qian, Y. & Sargent, A. C. *Sociol. Compass* **15**, e12881 (2021).
4. Cardel, M. I. *et al. J. Women Health.* **29**, 721–733 (2020).
5. Ward, H. B., Levin, F. R. & Greenfield, S. F. *Acad. Med.* **97**, 487–491 (2022).

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-022-01289-8>

- WHERE I WORK
- 10 May 2022

Immune to frustrations

Mercedes Segovia researches the role of dendritic cells in organizing the immune system.

- [Patricia Maia Noronha](#)



Mercedes Segovia is an immunologist at the Pasteur Institute of Montevideo. Credit: Pablo Albarenga for *Nature*

In this image, I am using a cell sorter to isolate dendritic cells — a type of blood cell — from mice with cancer. There are relatively few of these cells in the body, but they have a crucial role in the function of the immune

system. They incite the immune system to act against ‘invaders’ such as tumours, viruses and bacteria.

I studied medicine at the University of the Republic in Montevideo. In 2003, I moved to France with [my husband, Marcelo Hill](#), who was studying for a PhD in immunology in Nantes. I did my medical internship at the hospital there.

During that year, I became curious about what happens before patients reach the hospital: I wanted to help people before they start to feel ill, as opposed to just trying to treat symptoms. I started a master’s degree in immunology, and followed that with a PhD — also in immunology.

We stayed in France working as scientists for ten years — both of our children were born there. In 2013, however, we decided to return to Uruguay to share the knowledge we had acquired and to advance our careers in our home country. My husband got the chance to set up his own lab, at the Pasteur Institute of Montevideo, where I work as an associate researcher.

It’s very productive to have my husband as a colleague: our minds are used to working well together.

In the lab, the two of us, alongside eight others, are working on a project studying the protein TMEM176B. Our team showed that this protein keeps dendritic cells from triggering inflammation; inhibiting it could play a crucial part in the fight against cancer, and boosting it could help the process of organ transplantation.

I always look forward to seeing our findings adopted by the medical community, but it can take decades to happen, which can be annoying.

I think to be successful in this career, you must have passion, patience and — above all — a tolerance for frustration.

Nature **605**, 580 (2022)

doi: <https://doi.org/10.1038/d41586-022-01302-0>

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-022-01302-0>

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

Research

- **[A trail of dark-matter-free galaxies from a bullet-dwarf collision](#)** [18 May 2022]
Article • The dark-matter-free dwarf galaxies DF2 and DF4 in the NGC 1052 group probably formed together in the aftermath of a single bullet-dwarf collision around eight billion years ago.
- **[Direct observation of the dead-cone effect in quantum chromodynamics](#)** [18 May 2022]
Article • The direct measurement of the QCD dead cone in charm quark fragmentation is reported, using iterative declustering of jets tagged with a fully reconstructed charmed hadron.
- **[Microcomb-driven silicon photonic systems](#)** [18 May 2022]
Article • A simple and power-efficient microcomb source is used to drive complementary metal–oxide–semiconductor silicon photonic engines, a step towards the next generation of fully integrated photonic systems.
- **[Microfluidic chain reaction of structurally programmed capillary flow events](#)** [18 May 2022]
Article • Microfluidic chain reactions encode programs structurally in situ, and can form a frugal, versatile, bona fide lab-on-a-chip with wide-ranging applications in liquid handling and point-of-care diagnostics
- **[Synergistic HNO₃–H₂SO₄–NH₃ upper tropospheric particle formation](#)** [18 May 2022]
Article • By performing experiments under upper tropospheric conditions, nitric acid, sulfuric acid and ammonia can form particles synergistically, at rates orders of magnitude faster than any two of the three components.
- **[Genetic and chemotherapeutic influences on germline hypermutation](#)** [11 May 2022]
Article • A study of 21,879 families with rare genetic diseases identifies 12 with 2- to 7-fold excess of germline mutations, most of which are due to DNA repair defects or exposure to mutagenic chemotherapy, although most individuals with a hypermutated genome will not have a genetic disease.
- **[Phage anti-CBASS and anti-Pycsar nucleases subvert bacterial immunity](#)** [08 April 2022]

Article • A study using a biochemical screen of 57 phages in two bacterial species identifies and characterizes proteins enabling phages to evade CBASS and Pycsar immune systems, and describes the mechanisms involved.

- **Nonlinear mechanics of human mitotic chromosomes** [04

May 2022]

Article • A method that uses a combination of optical trapping, fluorescence microscopy and microfluidics to analyse the internal structure of chromosomes shows that there is a distinct nonlinear stiffening of the chromosome in response to tension.

- **Design of protein-binding proteins from the target structure alone** [24 March 2022]

Article • A design pipeline is presented whereby binding proteins can be designed de novo without the need for prior information on binding hotspots or fragments from structures of complexes with binding partners.

- **USP14-regulated allostery of the human proteasome by time-resolved cryo-EM** [27 April 2022]

Article • Structures of the human ubiquitin-specific protease 14 in complex with the 26S proteasome captured in the act of protein degradation provide a detailed view of the functional cycle of the USP14-regulated proteasome.

[Download PDF](#)

- Article
- Open Access
- [Published: 18 May 2022](#)

A trail of dark-matter-free galaxies from a bullet-dwarf collision

- [Pieter van Dokkum](#) ORCID: orcid.org/0000-0002-8282-9888¹,
- [Zili Shen](#) ORCID: orcid.org/0000-0002-5120-1684¹,
- [Michael A. Keim](#)¹,
- [Sebastian Trujillo-Gomez](#) ORCID: orcid.org/0000-0003-2482-0049²,
- [Shany Danieli](#)³,
- [Dhruba Dutta Chowdhury](#)¹,
- [Roberto Abraham](#)⁴,
- [Charlie Conroy](#) ORCID: orcid.org/0000-0002-1590-8551⁵,
- [J. M. Diederik Kruijssen](#) ORCID: orcid.org/0000-0002-8804-0212²,
- [Daisuke Nagai](#)⁶ &
- [Aaron Romanowsky](#) ORCID: orcid.org/0000-0003-2473-0369^{7,8}

Nature volume **605**, pages 435–439 (2022)

- 1372 Accesses
- 177 Altmetric
- [Metrics details](#)

Subjects

- [Dark energy and dark matter](#)
- [Galaxies and clusters](#)

Abstract

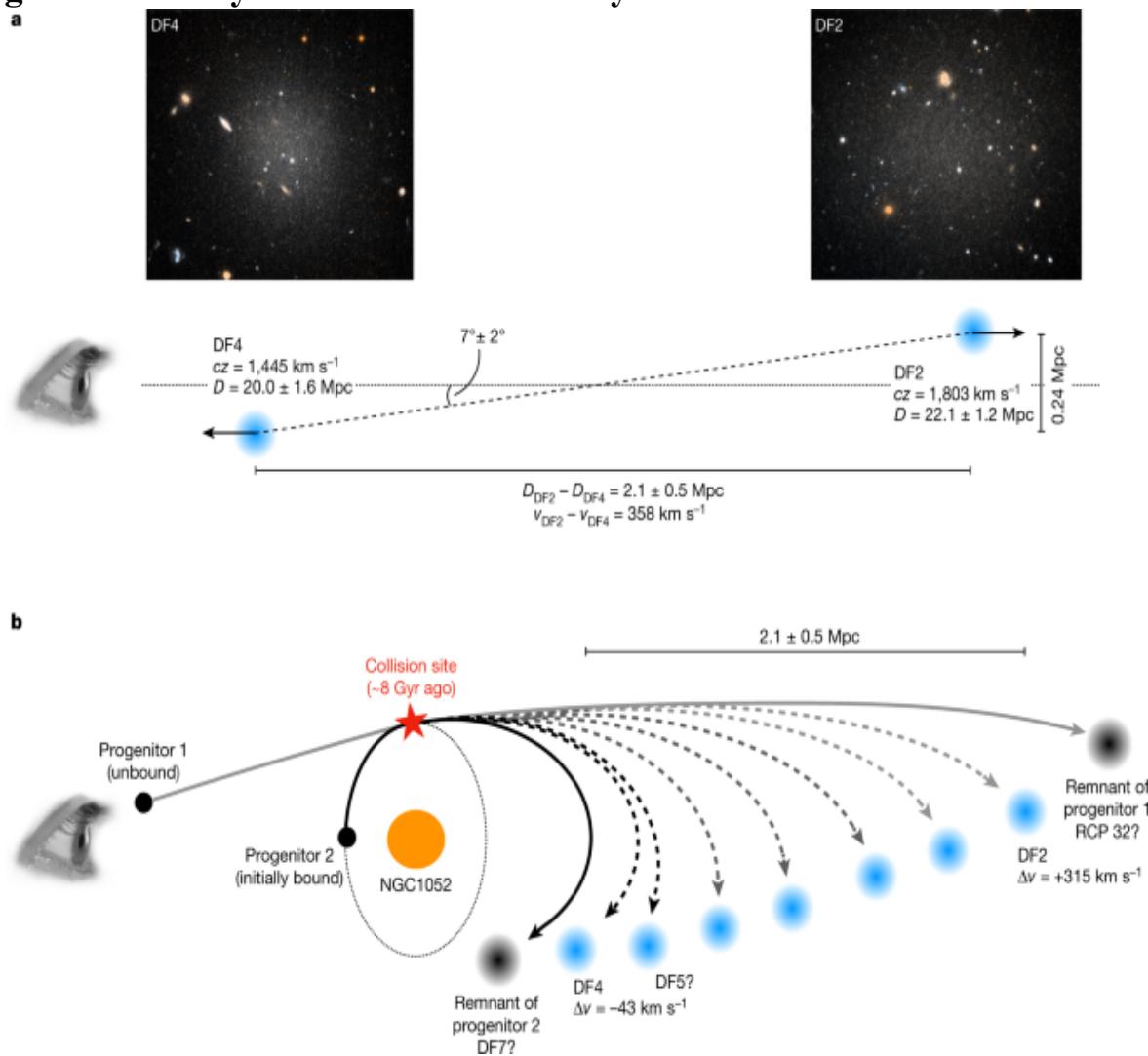
The ultra-diffuse galaxies DF2 and DF4 in the NGC 1052 group share several unusual properties: they both have large sizes¹, rich populations of overluminous and large globular clusters^{2,3,4,5,6}, and very low velocity dispersions that indicate little or no dark matter^{7,8,9,10}. It has been suggested that these galaxies were formed in the aftermath of high-velocity collisions of gas-rich galaxies^{11,12,13}, events that resemble the collision that created the bullet cluster¹⁴ but on much smaller scales. The gas separates from the dark matter in the collision and subsequent star formation leads to the formation of one or more dark-matter-free galaxies¹². Here we show that the present-day line-of-sight distances and radial velocities of DF2 and DF4 are consistent with their joint formation in the aftermath of a single bullet-dwarf collision, around eight billion years ago. Moreover, we find that DF2 and DF4 are part of an apparent linear substructure of seven to eleven large, low-luminosity objects. We propose that these all originated in the same event, forming a trail of dark-matter-free galaxies that is roughly more than two megaparsecs long and angled $7^\circ \pm 2^\circ$ from the line of sight. We also tentatively identify the highly dark-matter-dominated remnants of the two progenitor galaxies that are expected¹¹ at the leading edges of the trail.

Main

We begin with the assumption that it is not a coincidence that the ultra-diffuse galaxies DF2 and DF4 in the NGC 1052 group have the same set of otherwise-unique properties and that they were in close proximity to one another at the time of their formation. With that assumption, collisional formation is implied by their present-day radial velocities and three-dimensional locations. The geometry is shown in Fig. 1a. The relative radial velocity of the galaxies is high, 358 km s^{-1} , which is three times the velocity dispersion of the NGC 1052 group (about 115 km s^{-1}). Furthermore, although the two galaxies are separated by only 0.24 Mpc in the plane of the sky³, a differential tip of the red giant branch (TRGB) analysis has shown that they are $2.1 \pm 0.5 \text{ Mpc}$ apart along the line of sight^{15,16}, which is five times the virial radius of NGC 1052 (about 400 kpc; refs. [16,17](#)). In the

context of a shared origin of both galaxies, their radial velocities are consistent with their line-of-sight distances, that is, the closest galaxy (DF4) is moving towards us (with respect to the mean velocity) and the farthest galaxy (DF2) is moving away from us. Tracing their line-of-sight positions back in time, we infer that they must have formed in a high-velocity encounter. The minimum time since that encounter is about 6 Gyr, for constant motion along the line of sight.

Fig. 1: Geometry of the DF2 and DF4 system.



a, The radial velocity difference between DF2 and DF4 is 358 km s^{-1} and this large velocity difference is accompanied by a large line-of-sight separation of $2.1 \pm 0.5 \text{ Mpc}$ (refs. [15,16](#)). The geometry implies that the galaxies are moving away from each other. Tracing their positions back in

time, we infer that they were formed in a high-speed encounter ≥ 6 Gyr ago. **b**, Example of a collisional scenario involving NGC 1052. Velocities are given with respect to that galaxy ($cz = 1,488 \text{ km s}^{-1}$). An infalling gas-rich galaxy on an unbound orbit collided with a satellite of NGC 1052 about 8 Gyr ago, leading to two dark remnants (possibly RCP 32 and DF7), DF2 and DF4, and three to seven other dark-matter-free galaxies.

The situation at that point, a $\sim 300 \text{ km s}^{-1}$ encounter between two galaxies in the gas-rich environment of a young group, is a close match to the initial conditions of mini-bullet-cluster¹⁴ scenarios that have previously been proposed for the formation of DF2 and DF4^{11,12,13}. In a near head-on collision between two gas-rich galaxies, the collisional gas can be shocked and separated from the collisionless dark matter and pre-existing stars^{11,12}. The accompanying star formation favours massive clumps in highly compressed gas^{11,13}, producing the unusual globular clusters and the lack of dark matter. Galaxies that form this way are initially compact¹³, in apparent conflict with the large half-light radii of DF2 and DF4. However, intense feedback accompanying the formation of the globular clusters is expected to increase the sizes of the newly formed galaxies, an effect that is particularly efficient when the stars are not bound by a dark-matter halo^{18,19}. These previous studies focused on the formation of a single new galaxy at the collision site, but we propose that both DF2 and DF4 were formed in a single ‘bullet dwarf’ event. An illustration of the proposed scenario is shown in [Methods](#).

We have sufficient information to construct a plausible model for the geometry and timing of the collision. It is likely that the event took place near the central elliptical galaxy, NGC 1052, as it is roughly halfway between DF2 and DF4 in projection and its deep potential well is conducive to high-speed interactions. It is also likely that at least one of the progenitor galaxies was a satellite of NGC 1052 as the probability of a collision of two unbound galaxies is extremely small¹². We assume that the second progenitor was not bound to NGC 1052, and either on first infall or a satellite of another massive galaxy in the group²⁰. DF4 has a velocity difference of -43 km s^{-1} with respect to NGC 1052, whereas the velocity of DF2 is $+315 \text{ km s}^{-1}$. In Fig. [1b](#), we show a configuration for DF2, DF4 and NGC 1052 that satisfies these constraints. Progenitor 1 arrived in the vicinity

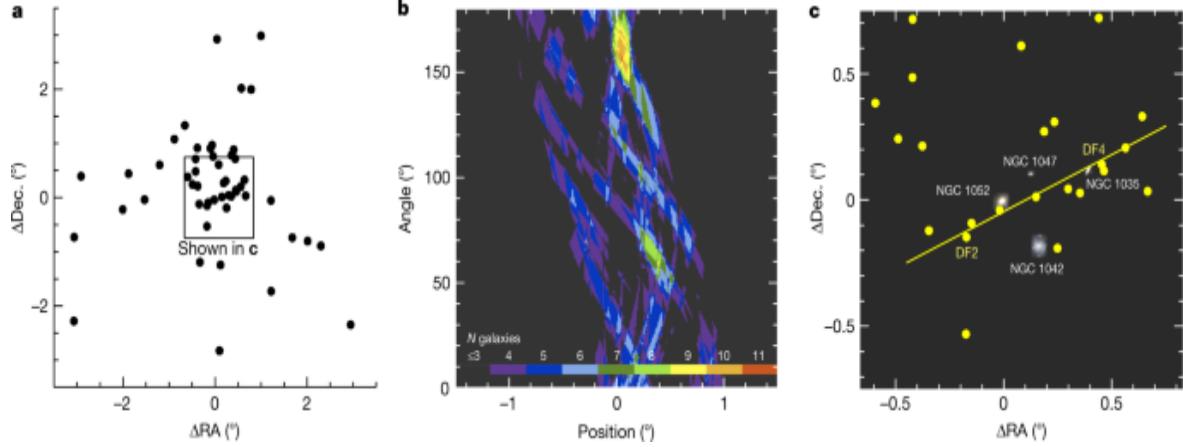
of NGC 1052 with a high ($>300 \text{ km s}^{-1}$) relative velocity and collided with progenitor 2, which was on a bound orbit. The collision produced DF2 and DF4, with DF2's velocity and orbit similar to progenitor 1 and DF4's similar to progenitor 2. DF2 is currently unbound and about 3 Mpc behind the group, whereas DF4 remained bound and has begun falling back. This geometry is not unique but it is similar to examples that have been explored in simulations¹². It is also consistent with the near-identical tidal distortions of the two galaxies²¹: in the geometry of Fig. 1b, DF2 and DF4 were at the same distance from NGC 1052 when they were formed, and as neither galaxy experienced a stronger tidal field afterwards, their morphologies have remained the same. In this model, DF2 has travelled about 3 Mpc since the collision with an average velocity with respect to NGC 1052 that is probably higher than its present-day value of 315 km s^{-1} . Assuming an average post-collision velocity of $\langle v \rangle \approx 350 \text{ km s}^{-1}$ dates the collision to about 8 Gyr ago, in excellent agreement with the ages of the globular clusters and the diffuse light in DF2 (9 ± 2 Gyr, as measured from optical spectra^{2,22}).

We further investigate the possible joint formation of DF2 and DF4 by examining the spatial distribution of galaxies along the DF2–DF4 axis. Given the complex gas distribution during and after the event, it may be that more than two dark-matter-deficient objects were formed in the wake of the collision¹². Furthermore, the bullet-dwarf scenario predicts the existence of two dark-matter-dominated objects that are the remnants of the progenitor galaxies. These should precede DF2, DF4 and any other dark-matter-deficient galaxies along their path, as they have the highest velocities relative to the barycentre¹¹.

The spatial distribution of galaxies with magnitude $g > 16.5$ in a recently compiled catalogue of the NGC 1052 field²³ is shown in Fig. 2a. An objective search for linear features in this distribution is performed using a discrete implementation of the Hough transform²⁴. The transform is shown in Fig. 2b. There is a clear peak with 11 galaxies on a line, corresponding to the relation

$$\begin{aligned} \text{\$}\{ \text{\rm\{ \bigtriangleup \}} \} \{ \text{\rm\{ D \}} \} \{ \text{\rm\{ e \}} \} \\ \{ \text{\rm\{ c \}} \} = 0.45 \{ \text{\rm\{ \bigtriangleup \}} \} \{ \text{\rm\{ R \}} \} \{ \text{\rm\{ A \}} \} - 2.8, \end{aligned} \quad (1)$$

Fig. 2: A linear feature in the spatial distribution of faint galaxies in the NGC 1052 field.



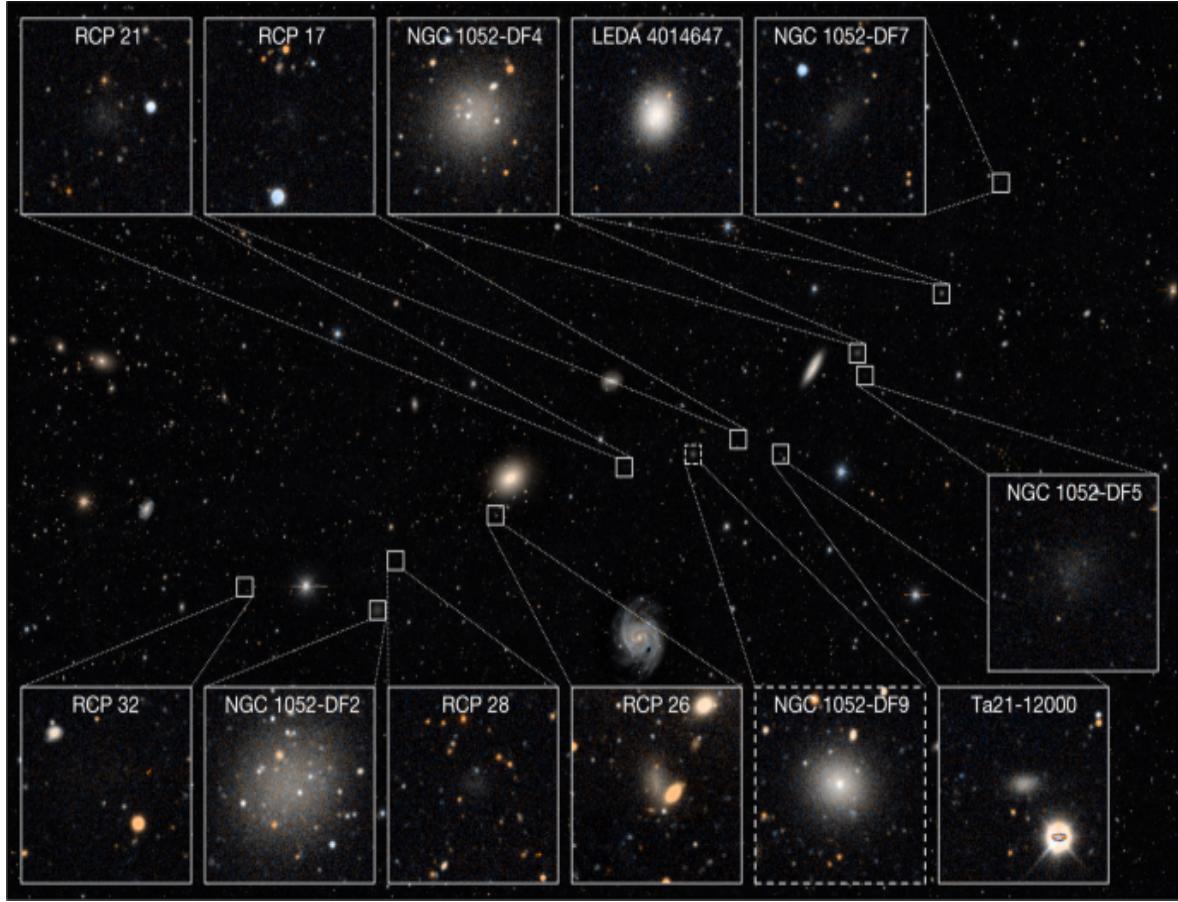
a, Distribution of galaxies with $g > 16.5$ (black circles) from a recent compilation of low-surface-brightness objects in the NGC 1052 field²³. The positions are with respect to the coordinates of NGC 1052, the central bright elliptical galaxy in the group. **b**, Hough transform²⁴ of the spatial distribution. The peak corresponds to a line that has 11 galaxies located within ± 30 kpc. The significance of the feature, as determined from randomized realizations of the data, is 97%. **c**, Zoom of **a**, shown with the four brightest members of the NGC 1052 group. Yellow circles correspond to the positions of the galaxies in the box shown in **a**. The orientation and offset of the line corresponds to the location of the peak of the Hough transform. Both DF2 and DF4 are part of the linear feature.

in units of arcminutes north and west of NGC 1052 (where Dec. is declination and RA is right ascension). This relation is shown by the line in Fig. 2c. Both DF2 and DF4 are in the sample of 11 galaxies. The probability that the peak arose by chance is 3%, and the probability that the peak arose by chance and that both DF2 and DF4 are part of it is 0.6% ([Methods](#)). Before turning to the properties of the galaxies in the trail, we note that the Hough transform provides post hoc validation of our initial assumption that DF2 and DF4 are related to each other.

Images of the 11 galaxies that are part of the trail are shown in Fig. 3. The average galaxy density in the central projected radius from NGC 1052 $R < 30'$ implies that 2 ± 2 of the 11 galaxies are chance projections, and we infer that there are 7–11 galaxies in the structure. Besides DF2 and DF4, other

galaxies in the trail are also unusually large for their luminosity. The relation between size and apparent magnitude for faint galaxies in the NGC 1052 group is shown in Fig. 4a. After subtracting a simple linear-least-squares fit to the running median (dashed line), we find that galaxies in the trail are on average 26% larger than the rest of the sample. The Wilcoxon probability that the trail galaxies and the rest of the sample are drawn from the same size distribution is <1%. The spatial distribution of galaxies colour-coded by their (magnitude dependent) relative size is shown in Fig. 4b. The unusual prevalence of large, low-surface-brightness galaxies in the central regions of the NGC 1052 group has been noted previously^{1,23}; here we propose that the bullet-dwarf event was responsible for it.

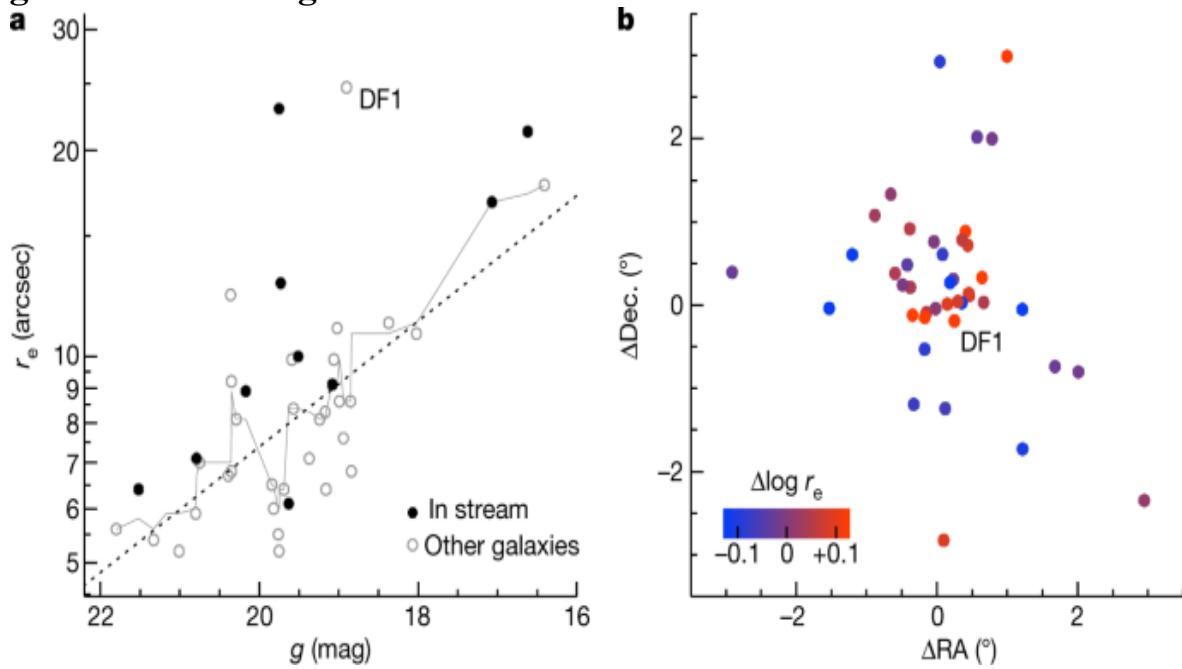
Fig. 3: Galaxies on the DF2–DF4 axis.



Legacy survey image of the central region of the NGC 1052 group, highlighting the 11 galaxies that are part of the trail according to the Hough transform. Several (2 ± 2) of these are expected to be chance projections. LEDA 4014647 is a candidate for an interloper (that is, an unrelated group

member) given its brightness and relative compactness. Its radial velocity was listed as $1,680 \pm 60$ in earlier SDSS releases (Data Release 3) but was later erroneously revised to a $z=0.7$ quasi-stellar object (Data Release 16). Judging from morphology alone, RCP 26 and Ta21-12000 may also be chance projections. Besides DF2 and DF4, RCP 32, DF5 and DF7 all satisfy or nearly satisfy the ultra-diffuse galaxy criteria. RCP 32 and DF7 are candidates for the two dark-matter-dominated remnants that have been predicted to precede dark-matter-deficient galaxies along the post-collision trajectory¹¹. We also highlight DF9 (SDSS J024007.01–081344.4), a galaxy with a bright star cluster that falls on the trail but is not part of the objectively selected sample²³.

Fig. 4: The sizes of galaxies on the DF2–DF4 axis.



a, The apparent size–apparent magnitude relation for low-surface-brightness objects in NGC 1052, using recent measurements from Legacy data²³. Galaxies that are part of the trail are labelled with solid symbols. The thin solid line is a running median, with $N=7$. The dashed line is a least-squares fit to the running median and has the form $\log r_e = -0.09(g - 20) + 0.87$ kpc. Galaxies in the trail are typically larger than other galaxies of the same magnitude. **b**, Distribution of galaxies colour-coded by their location with respect to the least-squares fit. The largest galaxies in the group are preferentially located along the DF2–DF4 axis.

We highlight three large and unusual galaxies that are part of the sample of 11. DF5 looks very similar to DF2 and DF4 but has a much lower surface brightness^{1,21}. It is so close in projection to DF4 (only 1.6' away) that both galaxies were observed in the same Hubble Space Telescope (HST) Advanced Camera for Surveys (ACS) pointing^{1,9}. In the bullet-dwarf scenario, this puzzling object is readily explained as another dark-matter-free fragment that formed in the aftermath of the collision. RCP 32 and DF7 are the farthest away from the centre of the structure, and are ‘ahead’ of DF2 and DF4, respectively. RCP 32 is an extremely faint ultra-diffuse galaxy and may have a globular cluster system²³. DF7 is an elongated ultra-diffuse galaxy that was previously observed with the HST¹ ([Methods](#)). We tentatively identify RCP 32 and DF7 as the candidate remnants of the original, pre-collision galaxies. In the geometry of Fig. [1b](#), RCP 32 could be the remnant of progenitor 1, the gas-rich object that was not bound to NGC 1052, and the brighter galaxy DF7 could be the remnant of progenitor 2, the galaxy that was a satellite of NGC 1052 at the time of collision. In this context, the elongation can be understood as the effect of the strong tidal forces at the time of collision²⁵.

The scenario that is proposed here makes predictions for the properties of the collision products, and further observations can test and refine this explanation. As the formation of DF2 and DF4 was triggered by a single event, the ages of the globular clusters of DF4, which have not yet been measured, should be identical to those of the clusters in DF2 (9 ± 2 Gyr; refs. [2,22](#)). A stringent and model-independent test is to directly compare the (averaged) spectra of clusters in the two galaxies; any clear differences could falsify our model, particularly differences in age-sensitive spectral features. Turning to other galaxies on the trail, their kinematics are predicted to be consistent with baryon-only models—except RCP 32 and DF7, which could show evidence for an unusually low baryon fraction. Furthermore, the radial velocities of trail galaxies should follow the approximate relation $cz \approx 1,700 - 10\Delta RA \text{ km s}^{-1}$, where ΔRA is defined as in equation ([1](#)), c is the speed of light and z is the redshift. Similarly, their line-of-sight distances are predicted to be $D \approx 21.5 - 0.06\Delta RA \text{ Mpc}$, for TRGB distances that are calibrated to $D = 22.1 \text{ Mpc}$ for DF2¹⁶. We note that these relations probably have considerable scatter and are not expected to be linear for most geometries. Also, 0–4 galaxies of the 11 are expected to be interlopers (Fig.

[3](#)), and some galaxies that seem to be off the trail may in fact be part of it due to the foreshortening (Fig. [1a](#)). This may be the case for DF1, a very large and diffuse galaxy (marked in Fig. [4](#)) that is only $14'$ off the DF2–DF4 axis and is elongated towards the centre of the trail.

Bullet-dwarf collisions hold the potential to constrain the self-interaction cross-section of dark matter. Modelling of the bullet cluster has provided an upper limit^{[26](#)}, but as self-interacting dark matter was introduced to explain the ‘cored’ dark-matter density profiles of low-mass galaxies^{[27](#)}, it is important to measure the cross-section on small scales^{[28](#)}. Quantitative constraints will probably require more than a single example of a bullet dwarf. Encouragingly, these events are probably more common than the collision that produced the bullet cluster^{[29](#)}; a search for plausible DF2 and DF4 progenitors in the IllustrisTNG cosmological simulation^{[30](#)} produced 248 head-on high velocity collisions in a 100^3-Mpc^3 volume^{[12](#)}, corresponding to about 8 within $D < 20 \text{ Mpc}$.

Methods

Illustration of the collision scenario

The proposed scenario for the formation of DF2, DF4 and the other trail galaxies is shown in Extended Data Fig. [1](#). As discussed in the main text, the scenario is a combination of the original idea that a bullet-dwarf collision might have formed DF2 and/or DF4^{[11](#)}; the results from subsequent hydrodynamical simulations, showing that multiple dark-matter-free clumps can form in such a collision^{[12](#)} and that the formation of massive star clusters is indeed promoted^{[13](#)}; and the independent finding that feedback from massive cluster formation in these conditions leads to a rapid expansion of the galaxies^{[18](#)}.

Faint galaxy sample

We make use of a recently compiled catalogue of low-surface-brightness objects in the NGC 1052 field^{[23](#)}, augmented by a catalogue of all brighter galaxies with redshifts in the range of $1,000 \text{ km s}^{-1} < cz < 2,000 \text{ km s}^{-1}$ that

is provided in the same study. Reference ²³ makes use of the publicly available Dark Energy Camera Legacy Survey (DECaLS) dataset³¹. The galaxies were initially identified with a combination of automated techniques and visual inspection, with the majority coming from visual inspection. Their structural parameters were measured with IMFIT³². We caution that the DECaLS dataset suffers from sky subtraction errors around low-surface-brightness galaxies, and that this may bias the size measurements. The main point of Fig. 4 is a relative comparison of the sizes of galaxies on and off the trail and this should be more robust than the absolute size measurements.

Velocity dispersion of the NGC 1052 group

We use the latest compilation of radial velocities in the NGC 1052 field²³ for an updated value of the velocity dispersion of the group. Table 2 of ref. ²³ contains 30 galaxies with redshifts $cz < 2,000 \text{ km s}^{-1}$. Two were removed: DF2, as it is almost certainly not bound to the group, and LEDA 4014647. LEDA 4014647 was assigned a radial velocity of $1,680 \pm 60 \text{ km s}^{-1}$ in earlier Sloan Digital Sky Survey (SDSS) releases (Data Release 3), but its redshift was later revised to $z = 0.7$ (Data Release 16). A visual inspection of the SDSS spectrum shows no clear features. Using the biweight estimator³³, we find a central velocity for the remaining 28 galaxies of $\langle cz \rangle = 1,435 \pm 20 \text{ km s}^{-1}$ and a line-of-sight velocity dispersion of $\sigma = 115 \pm 15 \text{ km s}^{-1}$.

The Hough transform

We use the Hough transform to look for linear features in the galaxy distribution, a standard method for detecting lines in images²⁴. The transform provides the number of galaxies along all possible directions, characterized by an angle and a distance from the centre. A width and maximum linear extent have to be chosen; we use $\pm 30 \text{ kpc}$ ($\pm 5.2'$) for the width and $< 400 \text{ kpc}$ ($69'$) for the linear extent. Although the exact number of galaxies that the Hough transform associates with the linear feature depends on the precise limits that are chosen, the qualitative results are not sensitive to them. In Fig. 2b, the orientation of the line is on the vertical axis and offset with respect to NGC 1052 on the horizontal axis.

Statistical significance of the trail

We use simulations to assess the probability that the alignment of the 11 galaxies arose by chance. We generate $N = 1,000$ realizations of the (x, y) pairs by maintaining the angular distance from NGC 1052 for each pair and randomizing the angle. This procedure ensures that the density profile of the sample is maintained for all realizations. We then create Hough transforms for all realizations and determine how often the strongest linear feature contains ≥ 11 galaxies. We find that the probability of a chance alignment of ≥ 11 galaxies is 3%.

This calculation assumes that galaxies are oriented randomly with respect to NGC 1052, and does not take into account anisotropy associated with the filamentary structure of the cosmic web^{34,35}. Galaxy groups are generally not spherical but have an average projected axis ratio of 0.77 (ref. 36). We examined the large-scale structure in the NGC 1052 field using a recently compiled catalogue of galaxies²³ in this general area. Selecting all low-surface-brightness galaxies that were identified in that study plus all bright galaxies with $cz < 2,000 \text{ km s}^{-1}$ gives a sample of 72 probable group members. Their distribution is shown in Extended Data Fig. 2. The smooth density field was calculated with the non-parametric kernel density estimator³⁷. There is no evidence for large-scale structure associated with the trail. In fact, there are no galaxies in the trail direction in the outskirts of the group, and the overall orientation of the group is perpendicular to the trail. The assumption of isotropy is therefore slightly conservative, in the sense that more galaxies will be scattered towards the line than away from it.

Finally, we note that the probability that there is a chance alignment and that it is a coincidence that both DF2 and DF4 are part of it is very low. This joint probability can be calculated directly for the isotropic case: of the 31 simulations that have ≥ 11 aligned galaxies only 6 have both DF2 and DF4 in the sample, corresponding to a combined probability of the observed arrangement of 0.6%.

A 12th low-surface-brightness dwarf galaxy on the trail

Visual inspection of the DECaLS imaging³¹ readily shows that there is a fairly prominent 12th galaxy that is part of the apparent trail. The object is SDSS J024007.01–081344.4 (ref. ³¹); it was previously pointed out as a likely low-luminosity group member with a central star cluster³⁸. It is not in the objective catalogue that we use for the main analysis²³. This may be because of its redshift in the SDSS database (it is erroneously listed as a $z=0.933$ active galactic nucleus) or because the light from the central cluster moved the object outside of the size and surface-brightness criteria. We refer to the galaxy as DF9 as that was the catalogue number in our initial Dragonfly catalogue¹. We do not use the galaxy in the objective analysis but we show its DECaLS image in Fig. 3. For convenience, we provide the coordinates of all trail galaxies in Extended Data Table 1.

HST imaging of the candidate dark galaxy DF7

DF7 is at one of the leading edges of the trail, ‘ahead’ of DF4. The galaxy was observed with HST/ACS as part of an exploratory survey of Dragonfly-identified low-surface-brightness galaxies in several groups¹. The observations constituted two orbits, one orbit in F606W and one orbit in F814W. In Extended Data Fig. 3, we show the HST imaging at two different contrast levels. The galaxy is elongated and appears distorted, with the elongation in the direction of DF4. DF7’s apparent distortion, combined with its location at the leading edge of the trail, lead us to speculate that the galaxy is the highly dark-matter-dominated remnant of one of the two progenitor galaxies. We note that DF7 may be largely disrupted in this interpretation: the observed¹ axis ratio is $b/a = 0.42$, but given the extreme foreshortening of the geometry the intrinsic axis ratio could be a stream-like approximately 1:20.

Other proposed scenarios

The joint formation of DF2 and DF4 in a bullet-dwarf event explains their lack of dark matter, large sizes, luminous and large globular clusters, striking similarity, large distance between them, large radial velocity difference, and the presence of a trail of other galaxies on the DF2–DF4 axis. Here we briefly discuss other scenarios that have been proposed to explain the properties of DF2 and DF4.

Initially, follow-up studies focused on possible errors in the measurements, either in the masses³⁹ or in the distances of the galaxies^{40,41}. However, with four independent velocity dispersion measurements^{3,8,9,10} (three for DF2 and one for DF4) and TRGB distances from extremely deep HST data^{15,16}, these issues have now largely been settled.

Most astrophysical explanations centre on the absence of dark matter only, and invoke some form of extreme tidal interaction (with NGC 1052 or other galaxies) to strip the dark matter (along with a large fraction of the initial stellar population)^{42,43,44,45}. These models do not explain the low metallicity of the galaxies, why there are two nearly identical objects in the same group, the newly discovered trail, or their overluminous and too-large globular clusters. The globular clusters, which have the same age (within the errors) as the diffuse light²², show that the galaxies were formed in an unusual way and did not merely evolve in an unusual way. Besides the bullet scenario, the only model that explains the globular clusters is a study of star formation in galaxies that are in the tails of the scatter in the halo mass–stellar mass relation^{18,19}. This model has ad hoc initial conditions and does not account for the presence of two near-identical galaxies, but the key aspects of it (the formation of luminous globular clusters in a compact configuration and the subsequent puffing up of the galaxies owing to feedback) probably apply to the collision products in the bullet scenario (see main text).

It has recently been suggested that DF2 and DF4 are entirely unrelated, with DF4 being stripped of its dark matter by NGC 1035, which is near it in projection, and DF2 a face-on disk galaxy with a normal dark-matter content^{46,47}. The association of DF4 with NGC 1035 is not seen in all datasets²¹, and there is no compelling evidence that DF2 is a disk⁷. Furthermore, the globular clusters and the trail remain unexplained, and there is the question of the likelihood that DF2 and DF4 have entirely different explanations but coincidentally share several otherwise-unique properties.

Data availability

The HST data for DF7 are available in the Mikulski Archive for Space Telescopes (MAST; <http://archive.stsci.edu>), under programme ID 14644. The Legacy Survey data shown in Fig. 3 are available at <https://www.legacysurvey.org/>. All other data that support the findings of this study are available in published studies that are referenced in the text.

Code availability

We have made use of standard data analysis tools in the Python environment.

References

1. Cohen, Y. et al. The Dragonfly Nearby Galaxies Survey. V. HST/ACS observations of 23 low surface brightness objects in the fields of NGC 1052, NGC 1084, M96, and NGC 4258. *Astrophys. J.* **868**, 96 (2018).
2. van Dokkum, P. et al. An enigmatic population of luminous globular clusters in a galaxy lacking dark matter. *Astrophys. J.* **856**, L30 (2018).
3. van Dokkum, P., Danieli, S., Abraham, R. & Romanowsky, A. J. A second galaxy missing dark matter in the NGC 1052 group. *Astrophys. J.* **874**, L5 (2019).
4. Dutta Chowdhury, D., van den Bosch, F. C. & van Dokkum, P. On the orbital decay of globular clusters in NGC 1052-DF2: testing a baryon-only mass model. *Astrophys. J.* **877**, 133 (2019).
5. Dutta Chowdhury, D., van den Bosch, F. C. & van Dokkum, P. On the evolution of the globular cluster system in NGC 1052-DF2: dynamical friction, globular-globular interactions, and galactic tides. *Astrophys. J.* **903**, 149 (2020).
6. Shen, Z., van Dokkum, P. & Danieli, S. A complex luminosity function for the anomalous globular clusters in NGC 1052-DF2 and NGC 1052-DF4. *Astrophys. J.* **909**, 179 (2021).

7. van Dokkum, P. et al. A galaxy lacking dark matter. *Nature* **555**, 629–632 (2018).
8. van Dokkum, P. et al. A revised velocity for the globular cluster GC-98 in the ultra diffuse galaxy NGC 1052-DF2. *Res. Not. Am. Astron. Soc.* **2**, 54 (2018).
9. Danieli, S., van Dokkum, P., Conroy, C., Abraham, R. & Romanowsky, A. J. Still missing dark matter: KCWI high-resolution stellar kinematics of NGC1052-DF2. *Astrophys. J.* **874**, L12 (2019).
10. Emsellem, E. et al. The ultra-diffuse galaxy NGC 1052-DF2 with MUSE. I. Kinematics of the stellar body. *Astron. Astrophys.* **625**, A76 (2019).
11. Silk, J. Ultra-diffuse galaxies without dark matter. *Mon. Not. R. Astron. Soc.* **488**, L24–L28 (2019).
12. Shin, E.-j et al. Dark matter deficient galaxies produced via high-velocity galaxy collisions in high-resolution numerical simulations. *Astrophys. J.* **899**, 25 (2020).
13. Lee, J., Shin, E.-j & Kim, J.-h Dark matter deficient galaxies and their member star clusters form simultaneously during high-velocity galaxy collisions in 1.25 pc resolution simulations. *Astrophys. J.* **917**, L15 (2021).
14. Clowe, D. et al. A direct empirical proof of the existence of dark matter. *Astrophys. J.* **648**, L109–L113 (2006).
15. Danieli, S. et al. A tip of the red giant branch distance to the dark matter deficient galaxy NGC1052-DF4 from deep Hubble Space Telescope data. *Astrophys. J.* **895**, L4 (2020).
16. Shen, Z. et al. A tip of the red giant branch distance of 22.1 ± 1.2 Mpc to the dark matter deficient galaxy NGC 1052-DF2 from 40 orbits of Hubble Space Telescope imaging. *Astrophys. J.* **914**, L12 (2021).

17. Forbes, D. A., Alabi, A., Brodie, J. P. & Romanowsky, A. J. Dark matter and no dark matter: on the halo mass of NGC 1052. *Mon. Not. R. Astron. Soc.* **489**, 3665–3669 (2019).
18. Trujillo-Gomez, S., Kruijssen, J. M. D., Keller, B. W. & Reina-Campos, M. Constraining the formation of NGC 1052-DF2 from its unusual globular cluster population. *Mon. Not. R. Astron. Soc.* **506**, 4841–4854 (2021).
19. Trujillo-Gomez, S., Kruijssen, J. M. D. & Reina-Campos, M. The emergence of dark matter-deficient ultra-diffuse galaxies driven by scatter in the stellar mass–halo mass relation and feedback from globular clusters. *Mon. Not. R. Astron. Soc.* **510**, 3356–3378 (2022).
20. van Dokkum, P., Danieli, S., Romanowsky, A., Abraham, R. & Conroy, C. The distance to NGC 1042 in the context of its proposed association with the dark matter-deficient galaxies NGC 1052-DF2 and NGC 1052-DF4. *Mon. Not. R. Astron. Soc.* **3**, 29 (2019).
21. Keim, M. A. et al. Tidal distortions in NGC1052-DF2 and NGC1052-DF4: independent evidence for a lack of dark matter. Preprint at <https://arxiv.org/abs/2109.09778> (2022).
22. Fensch, J. et al. The ultra-diffuse galaxy NGC 1052-DF2 with MUSE. II. The population of DF2: stars, clusters, and planetary nebulae. *Astron. Astrophys.* **625**, A77 (2019).
23. Román, J., Castilla, A. & Pascual-Granado, J. Discovery and analysis of low-surface-brightness galaxies in the environment of NGC 1052. *Astron. Astrophys.* **656**, A44 (2021).
24. Duda, R. O. & Hart, P. E. Use of the Hough transformation to detect lines and curves in pictures. *Commun. ACM* **15**, 11–15 (1972).
25. Rich, R. M. et al. A tidally distorted dwarf galaxy near NGC 4449. *Nature* **482**, 192–194 (2012).
26. Randall, S. W., Markevitch, M., Clowe, D., Gonzalez, A. H. & Bradač, M. Constraints on the self-interaction cross section of dark matter from

- numerical simulations of the merging galaxy cluster 1E 0657-56. *Astrophys. J.* **679**, 1173–1180 (2008).
27. Spergel, D. N. & Steinhardt, P. J. Observational evidence for self-interacting cold dark matter. *Phys. Rev. Lett.* **84**, 3760–3763 (2000).
 28. Tulin, S. & Yu, H.-B. Dark matter self-interactions and small scale structure. *Phys. Rep.* **730**, 1–57 (2018).
 29. Bouillot, V. R., Alimi, J.-M., Corasaniti, P.-S. & Rasera, Y. Probing dark energy models with extreme pairwise velocities of galaxy clusters from the DEUS-FUR simulations. *Mon. Not. R. Astron. Soc.* **450**, 145–159 (2015).
 30. Pillepich, A. et al. Simulating galaxy formation with the IllustrisTNG model. *Mon. Not. R. Astron. Soc.* **473**, 4077–4106 (2018).
 31. Dark Energy Survey Collaboration. The Dark Energy Survey: more than dark energy—an overview. *Mon. Not. R. Astron. Soc.* **460**, 1270–1299 (2016).
 32. Erwin, P. IMFIT: a fast, flexible new program for astronomical image fitting. *Astrophys. J.* **799**, 226 (2015).
 33. Beers, T. C., Flynn, K. & Gebhardt, K. Measures of location and scale for velocities in clusters of galaxies—a robust approach. *Astrophys. J.* **100**, 32–46 (1990).
 34. Bailin, J., Power, C., Norberg, P., Zaritsky, D. & Gibson, B. K. The anisotropic distribution of satellite galaxies. *Mon. Not. R. Astron. Soc.* **390**, 1133–1156 (2008).
 35. Tempel, E., Guo, Q., Kipper, R. & Libeskind, N. I. The alignment of satellite galaxies and cosmic filaments: observations and simulations. *Mon. Not. R. Astron. Soc.* **450**, 2727–2738 (2015).
 36. Wang, Y. et al. Probing the intrinsic shape and alignment of dark matter haloes using SDSS galaxy groups. *Mon. Not. R. Astron. Soc.* **385**, 1511–1522 (2008).

37. Silverman, B. W. *Density Estimation for Statistics and Data Analysis* (Chapman & Hall, 1986).
38. Trujillo, I. et al. Introducing the LBT Imaging of Galactic Halos and Tidal Structures (LIGHTS) survey. A preview of the low surface brightness Universe to be unveiled by LSST. *Astron. Astrophys.* **654**, A40 (2021).
39. Martin, N. F., Collins, M. L. M., Longeard, N. & Tollerud, E. Current velocity data on dwarf galaxy NGC 1052-DF2 do not constrain it to lack dark matter. *Astrophys. J.* **859**, L5 (2018).
40. Trujillo, I. et al. A distance of 13 Mpc resolves the claimed anomalies of the galaxy lacking dark matter. *Mon. Not. R. Astron. Soc.* **486**, 1192–1219 (2019).
41. Monelli, M. & Trujillo, I. The TRGB distance to the second galaxy “missing dark matter”: evidence for two groups of galaxies at 13.5 and 19 Mpc in the line of sight of NGC 1052. *Astrophys. J.* **880**, L11 (2019).
42. Ogiya, G. Tidal stripping as a possible origin of the ultra diffuse galaxy lacking dark matter. *Mon. Not. R. Astron. Soc.* **480**, L106–L110 (2018).
43. Macciò, A. V. et al. Creating a galaxy lacking dark matter in a dark matter-dominated universe. *Mon. Not. R. Astron. Soc.* **501**, 693–700 (2021).
44. Jackson, R. A. et al. Dark matter-deficient dwarf galaxies form via tidal stripping of dark matter in interactions with massive companions. *Mon. Not. R. Astron. Soc.* **502**, 1785–1796 (2021).
45. Moreno, J. et al. Galaxies lacking dark matter produced by close encounters in a cosmological simulation. *Nat. Astron.* <https://doi.org/10.1038/s41550-021-01598-4> (2022).
46. Montes, M. et al. The galaxy “missing dark matter” NGC 1052-DF4 is undergoing tidal disruption. *Astrophys. J.* **904**, 114 (2020).

47. Montes, M., Trujillo, I., Infante-Sainz, R., Monelli, M. & Borlaff, A. S. A disk and no signatures of tidal distortion in the galaxy “lacking” dark matter NGC 1052-DF2. *Astrophys. J.* **919**, 56 (2021).

Acknowledgements

A.R. was supported by National Science Foundation grant AST-1616710 and as a Research Corporation for Science Advancement Cottrell Scholar; J.M.D.K. acknowledges funding from the German Research Foundation (DFG) in the form of an Emmy Noether Research Group grant no. KR4801/1-1. J.M.D.K.; and S.T.-G. acknowledge funding from the European Research Council (ERC) under the European Union’s Horizon 2020 research and innovation programme via the ERC Starting Grant MUSTANG (grant agreement number 714907). Support from Space Telescope Science Institute grants HST-GO-14644, HST-GO-15695 and HST-GO-15851 is acknowledged. C. van Dokkum created Extended Data Fig. 1 and the eye photograph in Fig. 1.

Author information

Affiliations

1. Department of Astronomy, Yale University, New Haven, CT, USA
Pieter van Dokkum, Zili Shen, Michael A. Keim & Dhruba Dutta Chowdhury
2. Astronomisches Rechen-Institut, Zentrum für Astronomie der Universität Heidelberg, Heidelberg, Germany
Sebastian Trujillo-Gomez & J. M. Diederik Kruijssen
3. Department of Astrophysical Sciences, Princeton University, Princeton, NJ, USA
Shany Danieli

4. Department of Astronomy and Astrophysics, University of Toronto,
Toronto, Ontario, Canada

Roberto Abraham

5. Harvard-Smithsonian Center for Astrophysics, Cambridge, MA, USA

Charlie Conroy

6. Department of Physics, Yale University, New Haven, CT, USA

Daisuke Nagai

7. Department of Physics and Astronomy, San José State University, San
Jose, CA, USA

Aaron Romanowsky

8. University of California Observatories, Santa Cruz, CA, USA

Aaron Romanowsky

Contributions

P.v.D. led the analysis and wrote the manuscript. Z.S. performed the relative distance measurement (published in ref. [16](#)) that is at the basis of the study. M.A.K. created Fig. [3](#) and measured the structural parameters for DF5. S.T.-G. created an early version of Fig. [1b](#). All authors commented on the manuscript and aided in the interpretation.

Corresponding author

Correspondence to [Pieter van Dokkum](#).

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 Illustration of the proposed formation scenario of DF2 and DF4.](#)

Two gas-rich dwarf galaxies experience a high-speed encounter with a small impact parameter (top). Following previous studies^{11,12,13} the collisional gas gets stripped and shocked at closest approach, and forms stars at a prodigious rate with a bias towards massive clumps (second from top). The dark matter and previously formed stars are tidally distorted but continue ahead of the newly forming galaxies (third from top). Feedback in the absence of a dark-matter halo leads to expansion of the newly formed galaxies¹⁸. Most of the newly formed stellar mass is in two clumps, but several lower mass galaxies have also formed in the wake (bottom).

[Extended Data Fig. 2 Morphology of the NGC 1052 group.](#)

Distribution of 72 probable group members from a recent compilation of galaxies in the NGC 1052 field²³. Contours were derived with the non-parametric kernel density method³⁷. The dashed line indicates the trail.

There is no evidence that the trail is in the general direction of large-scale structure in this field.

Extended Data Fig. 3 A candidate dark-matter-dominated galaxy.

HST/ACS images¹ of DF7, which is located at the western end of the galaxy trail, beyond DF4. Left: a colour image generated from the F606W and F814W data. Right: a median-filtered version of the F606W image, with the arrow depicting the direction towards DF4. DF7 is a candidate for the dark-matter-dominated remnant of one of the original galaxies¹¹, given its location and its elongation in the direction of DF4.

Extended Data Table 1 Coordinates of the candidate trail galaxies

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 licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence 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 licence, visit <http://creativecommons.org/licenses/by/4.0/>.

Reprints and Permissions

About this article

Cite this article

van Dokkum, P., Shen, Z., Keim, M.A. *et al.* A trail of dark-matter-free galaxies from a bullet-dwarf collision. *Nature* **605**, 435–439 (2022).
<https://doi.org/10.1038/s41586-022-04665-6>

- Received: 29 November 2021
- Accepted: 21 March 2022
- Published: 18 May 2022
- Issue Date: 19 May 2022
- DOI: <https://doi.org/10.1038/s41586-022-04665-6>

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

[**How galaxies could exist without dark matter**](#)

- Nick Petrić Howe
- Noah Baker

Nature Nature Podcast 18 May 2022

[**Giant collision created galaxies devoid of dark matter**](#)

- Eun-jin Shin
- Ji-hoon Kim

Nature News & Views 18 May 2022

This article was downloaded by **calibre** from <https://www.nature.com/articles/s41586-022-04665-6>

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

[Download PDF](#)

- Article
- Open Access
- [Published: 18 May 2022](#)

Direct observation of the dead-cone effect in quantum chromodynamics

- [ALICE Collaboration](#)

Nature volume **605**, pages 440–446 (2022)

- 1849 Accesses
- 130 Altmetric
- [Metrics details](#)

Subjects

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

Abstract

In particle collider experiments, elementary particle interactions with large momentum transfer produce quarks and gluons (known as partons) whose evolution is governed by the strong force, as described by the theory of quantum chromodynamics (QCD)¹. These partons subsequently emit further partons in a process that can be described as a parton shower², which culminates in the formation of detectable hadrons. Studying the pattern of the parton shower is one of the key experimental tools for testing QCD. This

pattern is expected to depend on the mass of the initiating parton, through a phenomenon known as the dead-cone effect, which predicts a suppression of the gluon spectrum emitted by a heavy quark of mass m_Q and energy E , within a cone of angular size m_Q/E around the emitter³. Previously, a direct observation of the dead-cone effect in QCD had not been possible, owing to the challenge of reconstructing the cascading quarks and gluons from the experimentally accessible hadrons. We report the direct observation of the QCD dead cone by using new iterative declustering techniques^{4,5} to reconstruct the parton shower of charm quarks. This result confirms a fundamental feature of QCD. Furthermore, the measurement of a dead-cone angle constitutes a direct experimental observation of the non-zero mass of the charm quark, which is a fundamental constant in the standard model of particle physics.

Main

In particle colliders, quarks and gluons are produced in high-energy interactions through processes with large momentum transfer, which are calculable and well described by quantum chromodynamics (QCD). These partons undergo subsequent emissions, resulting in the production of more quarks and gluons. This evolution can be described in the collinear limit by a cascade process known as a parton shower, which transfers the original parton energy to multiple lower energy particles. This shower then evolves into a multi-particle final state, with the partons combining into a spray of experimentally detectable hadrons known as a jet⁶. The pattern of the parton shower is expected to depend on the mass of the emitting parton, through a phenomenon known as the dead-cone effect, whereby the radiation from an emitter of mass m and energy E is suppressed at angular scales smaller than m/E , relative to the direction of the emitter. The dead-cone effect is a fundamental feature of all gauge field theories (see ref. 3 for the derivation of the dead cone in QCD).

The dead-cone effect is expected to have sizeable implications for charm and beauty quarks, which have masses of $1.28 \pm 0.02 \text{ GeV}/c^2$ and $(4.18 \pm 0.02)^{+0.03} \text{ GeV}/c^2$ (ref. 1) in the minimal subtraction scheme, respectively, at energies on the GeV scale. The emission probability

in the collinear region, which is the divergent limit of QCD at which the radiation is most intense, is suppressed with increasing mass of the quark. This leads to a decrease in the mean number of particles produced in the parton shower. The DELPHI Collaboration at the LEP e^+e^- collider measured the multiplicity difference between events containing jets initiated by heavy beauty quarks and those containing light quarks (up, down or strange). They found that the differences depend only on the quark mass⁷, which was attributed to the suppression of collinear gluon radiation from the heavy quark because of the dead-cone effect. A measurement of the momentum density of jet constituents as a function of distance from the jet axis was also performed by the ATLAS collaboration at CERN⁸, which pointed to a depletion of momentum close to the jet axis that was ascribed as a consequence of the dead-cone effect. The mass of the beauty quark was also estimated through a phenomenological fit to the measured data⁹. As hard (large transverse momentum) emissions are preferentially emitted at small angles, and are therefore suppressed for massive emitters, heavy quarks also retain a larger fraction of their original momentum compared to lighter quarks, leading to a phenomenon known as the leading-particle effect. This has been well established experimentally, with the fraction of the jet momentum carried by the leading (highest transverse momentum) hadron containing a charm or beauty quark (heavy-flavour hadron) in jets, peaking at 0.6–0.7 and 0.8–0.9, respectively, whereas the corresponding fraction carried by the leading hadron in light quark-initiated jets peaks at smaller values^{10,11,12,13,14}.

Until now, a direct experimental measurement of the dead-cone effect has been subject to two main challenges. First, the dead-cone angular region can receive contributions from hadronization effects or particles that do not originate from the gluon radiation from the heavy-flavour quark, such as the decay products of heavy-flavour hadrons. The second difficulty lies in the accurate determination of the dynamically evolving direction of the heavy-flavour quark, relative to which the radiation is suppressed, throughout the shower process. The development of new experimental declustering techniques⁴ enables these aforementioned difficulties to be overcome by reconstructing the evolution of the jet shower, giving access to the kinematic properties of each individual emission. These techniques reorganize the particle constituents of an experimentally reconstructed jet, to access the

building blocks of the shower and trace back the cascade process. Isolated elements of the reconstructed parton shower that are likely to be unmodified by hadronization processes provide a good proxy for real quark and gluon emissions (splittings). These reclustering techniques have been demonstrated in inclusive (without tagging the initiating parton flavour) jets to successfully reconstruct splittings that are connected to or that preserve the memory of the parton branchings. This is demonstrated by measurements such as the groomed momentum balance^{15,16,17,18}, which probes the Dokshitzer–Gribov–Lipatov–Altarelli–Parisi splitting function¹⁹, and the Lund plane²⁰, which exposes the running of the strong coupling with the scale of the splittings. An experimental method to expose the dead cone in boosted top-quark events was also proposed in ref. ²¹.

Reclustering techniques are extended in this work to jets containing a charm quark based on the prescription given in ref. ²². These jets are tagged through the presence of a reconstructed D^0 meson amongst their constituents, which has a mass of $1.86 \text{ GeV}/c^2$ (ref. ¹) and is composed of a heavy charm quark and a light anti-up quark. The measurement is performed in proton–proton collisions at a centre-of-mass energy of $\sqrt{s} = 13 \text{ TeV}$ at the Large Hadron Collider (LHC), using the ALICE (A Large Ion Collider Experiment) detector. Further details of the detector apparatus and data measured can be found in the Methods. As the charm-quark flavour is conserved through the shower process, this provides an opportunity to isolate and trace back the emission history of the charm quark. In this way, by comparing the emission patterns of charm quarks to those of light quarks and gluons, the QCD dead cone can be directly revealed.

Selecting jets containing a D^0 meson

To select jets initiated by a charm quark, through the presence of a D^0 meson in their list of constituents, the D^0 mesons and jets need to be reconstructed in the events. The D^0 -meson candidates (and their anti-particles) were reconstructed in the transverse-momentum interval $(2 < \{p_T\} < 36) \text{ GeV}/c$, through the $D^0 \rightarrow K^- \pi^+$ (and charged conjugate) hadronic decay channel, which has a branching ratio of $3.95 \pm 0.03\%$ (ref. ¹). The D^0 -meson candidates were identified by

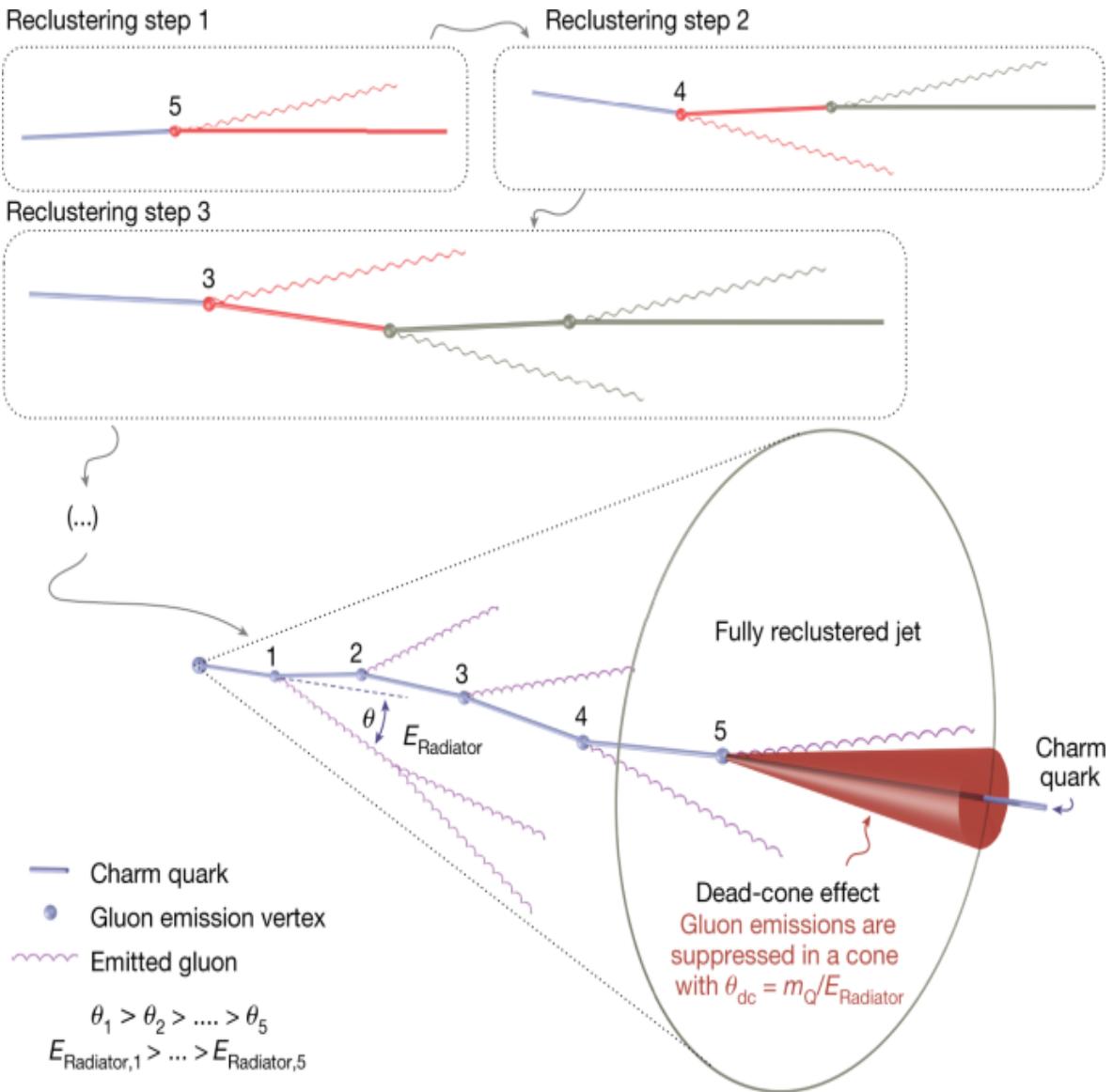
topological selections based on the displacement of the D^0 -meson candidate decay vertex, in addition to applying particle identification on the D^0 -meson candidate decay particles. These selection criteria largely suppress the combinatorial background of $K^\mp\pi^\pm$ pairs that do not originate from the decay of a D^0 meson. Further details on the selection criteria are provided in ref. [23](#).

Tracks (reconstructed charged-particle trajectories) corresponding to the D^0 -meson candidate decay particles were replaced by the reconstructed D^0 -meson candidate in the event, with the D^0 -meson candidate four-momentum being the sum of the decay-particle four-momenta. One benefit of this procedure is to avoid the case in which the decay products of the D^0 -meson candidate fill the dead-cone region. A jet-finding algorithm was then used to cluster the particles (tracks and the D^0 -meson candidate) in the event, to reconstruct the parton shower by sequentially recombining the shower particles into a single object (the jet). The jet containing the D^0 -meson candidate was then selected. The four-momentum of the jet is a proxy for the four-momentum of the charm quark initiating the parton shower. The jet-finding algorithm used was the anti- k_T algorithm [24](#) from the Fastjet package [25](#), which is a standard choice for jet reconstruction because of its high performance in reconstructing the original parton kinematics. More details on the jet finding procedure can be found in the Methods.

Reconstructing the jet shower

Once jets containing a D^0 -meson candidate amongst their constituents are selected, the internal cascade process is reconstructed. This is done by reorganizing (reclustering) the jet constituents according to the Cambridge–Aachen (C/A) algorithm [26](#), which clusters these constituents based solely on their angular distance from one another. A pictorial representation of this reclustering process, which starts by reconstructing the smallest angle splittings, is shown in the top panels of Fig. [1](#). As QCD emissions approximately follow an angular-ordered structure [27](#), the C/A algorithm was chosen as it also returns an angular-ordered splitting tree.

Fig. 1: Reconstruction of the showering quark.



A sketch detailing the reconstruction of the showering charm quark, using iterative declustering, is presented. The top panels show the initial reclustering procedure with the C/A algorithm, in which the particles separated by the smallest angles are brought together first. Once the reclustering is complete, the declustering procedure is carried out by unwinding the reclustering history. Each splitting node is numbered according to the declustering step in which it is reconstructed. With each splitting, the charm-quark energy, $E_{\text{Radiator},n}$, is reduced and the gluon is emitted at a smaller angle, θ_n , with respect to previous emissions. The mass of the heavy quark, m_Q , remains constant throughout the showering process. At each splitting, gluon emissions are suppressed in the dead-cone region

(shown by a red cone for the last splitting), which increases in angle as the quark energy decreases throughout the shower.

This splitting tree is then iteratively declustered by unwinding the reclustering history, to access the building blocks of the reconstructed jet shower. At each declustering step, two prongs corresponding to a splitting are returned. The angle between these splitting daughter prongs, θ , the relative transverse momentum of the splitting, k_T , and the sum of the energy of the two prongs, E_{Radiator} , are registered. As the charm flavour is conserved throughout the showering process, the full reconstruction of the D^0 -meson candidate enables the isolation of the emissions of the charm quark in the parton shower, by following the daughter prong containing the fully reconstructed D^0 -meson candidate at each declustering step. This can be seen in the bottom part of Fig. 1, which shows the evolution of the charm quark reconstructed from the measured final state particles. Moreover, the kinematic properties of the charm quark are updated along the splitting tree, enabling an accurate reconstruction of each emission angle against the dynamically evolving charm-quark direction. It was verified that in more than 99% of the cases the prong containing the D^0 -meson candidate at each splitting coincided with the leading prong. This means that following the D^0 -meson candidate or leading prong at each step is equivalent, and therefore a complementary measurement for an inclusive jet sample, when no flavour tagging is available, can be made by following the leading prong through the reclustering history. As the inclusive sample is dominated by massless gluon and nearly massless light quark-initiated jets, it acts as a reference to highlight the mass effects present in the charm tagged sample.

Extracting the true charm splittings

The selected sample of splittings has contributions from jets tagged with combinatorial $K^\mp\pi^\pm$ pairs, which are not rejected by the applied topological and particle identification selections. The measured invariant mass of real D^0 mesons, which corresponds to the rest mass, is distributed in a Gaussian (because of uncertainties in the measurement of the momenta of the $K^\mp\pi^\pm$ pairs) with a peak at the true D^0 -meson mass. This enables the implementation of a statistical two-dimensional side-band subtraction

procedure, which characterizes the background distribution of splittings by sampling the background-dominated regions of the D^0 -meson candidate invariant mass distributions, far away from the signal peak. In this way the combinatorial contribution can be accounted for and removed. Furthermore, the selections on the D^0 -meson candidates also select a fraction of D^0 mesons originating as a product of beauty-hadron decays. These were found to contribute 10–15% of the reconstructed splittings, with only a small influence on the results, which will be discussed later. The studies were performed using Monte Carlo (MC) PYTHIA 6.425 (Perugia 2011)^{28,29} simulations (this generator includes mass effects in the parton shower³⁰ and was used for all MC-based corrections in this work), propagating the generated particles through a detailed description of the ALICE detector with GEANT3 (ref. ³¹). The finite efficiency of selecting real D^0 -meson tagged jets, through the chosen selection criteria on the D^0 -meson candidates, as well as kinematic selections on the jets, was studied and accounted for through MC simulations. This efficiency was found to be strongly $\langle p_T \rangle \langle \ln D \rangle^0$ dependent and different for D^0 mesons originating from the hadronization of charm quarks or from the decay of beauty hadrons. Further details on these analysis steps can be found in the Methods.

As the reconstructed jet shower is built from experimentally detectable hadrons, as opposed to partons, hadronization effects must be accounted for. As hadronization processes occur at low non-perturbative scales, they are expected to distort the parton shower by mainly adding low- k_T splittings³². A selection of $k_T > 200 \text{ MeV}/c$ ensures that only sufficiently hard splittings are accepted and is used to suppress such hadronization effects. Other choices of k_T selection were also explored, with stronger k_T selections further removing non-perturbative effects from the measurement, at the expense of statistical precision. Other non-perturbative effects, such as the underlying event, contribute with extra soft splittings primarily at large angles and do not affect the small-angle region under study.

Detector effects also distort the reconstructed parton shower through inefficiencies and irresolution in the tracking of charged particles. However, these have been tested and largely cancel in the final observable, and any

residual effects are quantified in a data-driven way and included in the systematic uncertainties.

It should be noted that in addition to direct heavy-flavour pair creation in the elementary hard scattering, charm quarks can also be produced in higher-order processes as a result of gluon splitting. Therefore, the shower history of D^0 mesons containing such charm quarks will also have contributions from splittings originating from gluons. Furthermore, in the case of high transverse momentum gluons in which the charm quarks are produced close in angle to each other, the dead-cone region of the charm quark hadronizing into the reconstructed D^0 meson can be populated by particles produced in the shower, hadronization and subsequent decays of the other (anti-)charm quark. The influence of such contaminations through gluon splittings was studied with MC simulations and found to be negligible.

The observable $R(\theta)$

The observable used to reveal the dead cone is built by constructing the ratio of the splitting angle (θ) distributions for D^0 -meson tagged jets and inclusive jets, in bins of E_{Radiator} . This is given by

$$\begin{aligned} \$\$ R(\theta) = & \frac{N^{\text{D}}}{N^{\text{jets}}} \frac{\ln(1/\theta)}{\ln(N^{\text{inclusive}}/N^{\text{jets}})} \\ & \frac{\ln(d)}{\ln(k_T/T)} , \\ E_{\text{Radiator}} \end{aligned} \quad (1)$$

where the θ distributions were normalized to the number of jets that contain at least one splitting in the given E_{Radiator} and k_T selection, denoted by $(N^{\text{D}})^0/N^{\text{jets}}$ and $N^{\text{inclusive jets}}$ for the D^0 -meson tagged and inclusive jet samples, respectively. Expressing equation (1) in terms of the logarithm of the inverse of the angle is natural, given that at leading order the QCD probability for a parton to split is proportional to $\ln(1/\theta)\ln(k_T/T)$.

A selection on the transverse momentum of the leading track in the leading prong of each registered splitting in the inclusive jet sample, $\langle p \rangle_{\rm T, inclusive}^{\rm jets} \geq 2.8$ GeV/c, was applied. This corresponds to the transverse mass (obtained through the quadrature sum of the rest mass and transverse momentum) of a 2 GeV/c D⁰ meson and accounts for the $\langle p \rangle_{\rm T, inclusive}^{\rm jets} \geq 0$ selection in the D⁰-meson tagged jet sample, enabling a fair comparison of the two samples.

In the absence of mass effects, the charm quark is expected to have the same radiating properties as a light quark. In this limit, equation (1) can be rewritten as

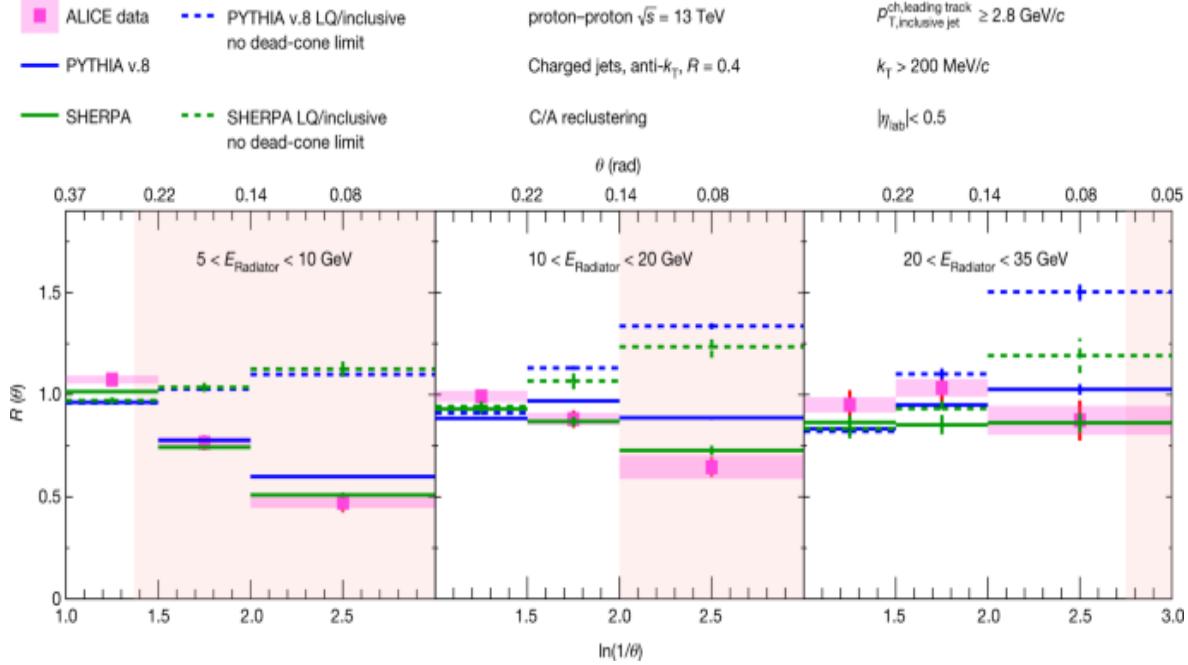
$$\begin{aligned} \$\$ R(\theta)_{\text{no dead-cone}} &= \frac{N_{\text{LQ}}^{\text{jets}}}{N_{\text{inclusive}}^{\text{jets}}} \frac{\ln(1/\theta)}{\ln(1/\theta_{\text{dead-cone}})} \\ &= \frac{N_{\text{LQ}}^{\text{jets}}}{N_{\text{inclusive}}^{\text{jets}}} \frac{\ln(d)}{\ln(d_{\text{dead-cone}})} \\ &= \frac{k_{\text{T}}}{E_{\text{Radiator}}} , \end{aligned} \quad (2)$$

where the superscript LQ refers to light quarks, and the inclusive sample contains both light-quark and gluon-initiated jets. This indicates that the $R(\theta)_{\text{no dead-cone}}$ ratio depends on the differences between light-quark and gluon radiation patterns, which originate from the fact that gluons carry two colour charges (the charge responsible for strong interactions) whereas quarks only carry one. These differences result in quarks fragmenting at a lower rate and more collinearly than gluons. Therefore, in the limit of having no dead-cone effect, the ratio of the θ distributions for D⁰-meson tagged jets and inclusive jets becomes $R(\theta)_{\text{no dead-cone}} > 1$, at small angles. This was verified through SHERPA v.2.2.8 (ref. 33) and PYTHIA v.8.230 (Tune 4C)³⁴ MC generator calculations, with the specific $R(\theta)_{\text{no dead-cone}}$ value dependent on the quark and gluon fractions in the inclusive sample. SHERPA and PYTHIA are two MC generators commonly used in high-energy particle physics and they use different shower prescriptions and hadronization models. Both models implement the dead-cone effect.

Exposing the dead cone

The measurements of $R(\theta)$, in the three radiator (charm-quark) energy intervals $5 < E_{\text{Radiator}} < 10 \text{ GeV}$, $10 < E_{\text{Radiator}} < 20 \text{ GeV}$ and $20 < E_{\text{Radiator}} < 35 \text{ GeV}$, are presented in Fig. 2. Detector effects largely cancel out in the ratio and results are compared to particle-level simulations. Residual detector effects are considered in the systematic uncertainty together with uncertainties associated with the reconstruction and signal extraction of D^0 -meson tagged jets, as well as detector inefficiencies in the reconstruction of charged tracks in both the D^0 -meson tagged and inclusive jet samples. More details on the study of systematic uncertainties can be found in the Methods.

Fig. 2: Ratios of splitting angle probability distributions.



The ratios of the splitting-angle probability distributions for D^0 -meson tagged jets to inclusive jets, $R(\theta)$, measured in proton–proton collisions at $\sqrt{s} = 13 \text{ TeV}$, are shown for $5 < E_{\text{Radiator}} < 10 \text{ GeV}$ (left panel), $10 < E_{\text{Radiator}} < 20 \text{ GeV}$ (middle panel) and $20 < E_{\text{Radiator}} < 35 \text{ GeV}$ (right panel). The data are compared with PYTHIA v.8 and SHERPA simulations, including the no dead-cone limit given by the ratio of the angular distributions for light-quark jets (LQ) to inclusive jets. The pink shaded

areas correspond to the angles within which emissions are suppressed by the dead-cone effect, assuming a charm-quark mass of $1.275 \text{ GeV}/c^2$.

A significant suppression in the rate of small-angle splittings is observed in D^0 -meson tagged jets relative to the inclusive jet population. In Fig. 2, the data are compared with particle-level SHERPA (green) and PYTHIA v.8.230 (blue) MC calculations, with SHERPA v.2.2.8 providing a better agreement with the data. The no dead-cone baseline, as described in equation (2), is also provided for each MC generator (dashed lines). The suppression of the measured data points relative to the no dead-cone limit directly reveals the dead cone within which the charm-quark emissions are suppressed. The coloured regions in the plots correspond to the dead-cone angles in each E_{Radiator} interval, $\theta_{\text{dc}} < m_Q/E_{\text{Radiator}}$, where emissions are suppressed. For a charm-quark mass $m_Q = 1.275 \text{ GeV}/c^2$ (ref. 1), these angles correspond to $\ln(1/\theta_{\text{dc}}) \geq 1.37, 2$ and 2.75 for the intervals $5 < E_{\text{Radiator}} < 10 \text{ GeV}$, $10 < E_{\text{Radiator}} < 20 \text{ GeV}$ and $20 < E_{\text{Radiator}} < 35 \text{ GeV}$, respectively. These values are in qualitative agreement with the angles at which the data start to show suppression relative to the MC limits for no dead-cone effect. The magnitude of this suppression increases with decreasing radiator energy, as expected from the inverse dependence of the dead-cone angle on the energy of the radiator.

A lower limit for the significance of the small-angle suppression is estimated by comparing the measured data to $R(\theta) = 1$, which represents the limit of no dead-cone effect in the case in which the inclusive sample is entirely composed of light quark-initiated jets. To test the compatibility of the measured data with the $R(\theta) = 1$ limit, a statistical test was performed by generating pseudodata distributions consistent with the statistical and systematic uncertainties of the measured data. A chi-square test was then carried out against this hypothesis for each of the pseudodata distributions. The mean P values correspond to significances of $7.7\sigma, 3.5\sigma$ and 1.0σ , for the $5 < E_{\text{Radiator}} < 10 \text{ GeV}$, $10 < E_{\text{Radiator}} < 20 \text{ GeV}$ and $20 < E_{\text{Radiator}} < 35 \text{ GeV}$ intervals, respectively. A σ value greater than 5 is considered the criteria for a definitive observation, whereas the value of 1.0 is consistent with the null hypothesis.

The MC distributions shown were generated separately for prompt (charm-quark initiated) and non-prompt (beauty-quark initiated) D^0 -meson tagged jet production and were then combined using the prompt and non-prompt fractions in data calculated with POWHEG³⁵ + PYTHIA v.6.425³⁴ simulations. The non-prompt fraction was found to be independent of the splitting angle and corresponds to approximately 10% of the splittings in the $5 < E_{\text{Radiator}} < 10$ GeV interval and approximately 15% of the splittings in both the $10 < E_{\text{Radiator}} < 20$ GeV and $20 < E_{\text{Radiator}} < 35$ GeV intervals. It was verified through the MC simulations that non-prompt D^0 -meson tagged jets should exhibit a smaller suppression at small angles in $R(\theta)$ compared with inclusive jets than their prompt counterparts. This is due to the additional decay products accompanying non-prompt D^0 -meson tagged jets that are produced in the decay of the beauty hadron. These may populate the dead-cone region, leading to a smaller observed suppression in $R(\theta)$, despite the larger dead-cone angle of the heavier beauty quark.

Conclusions

We have reported the direct measurement of the QCD dead cone, using iterative declustering of jets tagged with a fully reconstructed charmed hadron. The dead cone is a fundamental phenomenon in QCD, dictated by the non-zero quark masses, whose direct experimental observation has previously remained elusive. This measurement provides insight into the influence of mass effects on jet properties and provides constraints for MC models. These results pave the way for a study of the mass dependence of the dead-cone effect, by measuring the dead cone of beauty jets tagged with a reconstructed beauty hadron.

A future study of the dead-cone effect in heavy-ion collisions, in which partons interact strongly with the hot QCD medium that is formed and undergo energy loss through (dominantly) medium-induced radiation, is also envisaged. If a dead cone were observed for these medium-induced emissions, it would be a confirmation of the theoretical understanding of in-medium QCD radiation, which is a primary tool used to characterize the high-temperature phase of QCD matter^{36,37,38}.

The quark masses are fundamental constants of the standard model of particle physics and needed for all numerical calculations within its framework. Because of confinement, their values are commonly inferred through their influence on hadronic observables. An exception is the top quark, which decays before it can hadronize, as its mass can be constrained experimentally from the direct reconstruction of the decay final states³⁹ (see ref. ⁴⁰ for a review of top mass measurements at the Fermilab Tevatron and CERN LHC).

By accessing the kinematics of the showering charm quark, before hadronization, and directly uncovering the QCD dead-cone effect, our measurement provides direct sensitivity to the mass of quasi-free charm quarks, before they bind into hadrons.

Furthermore, future high-precision measurements using this technique on charm and beauty tagged jets, potentially in conjunction with machine-learning tools to separate quark and gluon emissions, could experimentally constrain the magnitude of the quark masses.

Methods

Detector setup and data set

The analysis was performed with the ALICE detector at the CERN LHC⁴¹. The ALICE Inner Tracking System⁴² and Time Projection Chamber⁴³ were used for charged-particle reconstruction, and particle identification (PID) was obtained using the combined information from the Time Projection Chamber and the Time-Of-Flight detectors⁴⁴. These detectors are located in the ALICE central barrel, which has full azimuthal coverage and a pseudorapidity range of $|\eta| < 0.9$. The data set used in this analysis was collected in 2016, 2017 and 2018 in proton–proton collisions at $\sqrt{s} = 13\text{ TeV}$, with a minimum-bias trigger condition defined by the presence of at least one hit in each of the two V0 scintillators⁴⁵. This trigger accepts all events of interest for this analysis and the collected data sample corresponds to an integrated luminosity of $\mathcal{L}_{\text{int}} = 25 \text{ nb}^{-1}$.

Jet finding and tagging

Jet finding was performed using the anti- k_T algorithm, with a jet resolution parameter of $R = 0.4$. The E -scheme recombination strategy was chosen to combine the tracks of the jet by adding their four-momenta, with a geometric constraint on the pseudorapidity of $|\eta| < 0.5$ enforced on the jet axis, to ensure that the full jet cone was contained in the acceptance of the central barrel of the ALICE detector. The ALICE detector has excellent tracking efficiency down to low p_T (approximately 80% at $p_T = 500 \text{ MeV}/c$), which is homogeneous as a function of pseudorapidity and azimuthal angle⁴¹, within the acceptance. The effect of track density on the tracking efficiency is also negligible⁴⁶. The angular resolution is about 20% down to splitting angles of 0.05 radians, which motivated a track-based jet measurement as opposed to a full jet measurement using calorimetric information. Recent measurements^{15,20} have shown that track-based jet observables are successful at reconstructing the parton shower information through declustering techniques, despite missing the information from the neutral component of the jet.

Jets with a transverse momentum in the interval of $\langle p_T \rangle \in [5, 15] \text{ GeV}/c$ were selected for this analysis. To mitigate against the cases in which two D^0 -meson candidates share a common decay track, jet-finding passes were performed independently for each D^0 -meson candidate in the event, each time replacing only the decay tracks of that candidate with the corresponding D^0 -meson candidate. In each pass the jet containing the reconstructed D^0 -meson candidate of that pass was subsequently tagged as a charm-initiated jet candidate.

Subtraction of the combinatorial background in the D^0 -meson candidate sample

To extract the true D^0 -meson tagged jet $R(\theta)$ distributions and remove the contribution from combinatorial $K^\mp\pi^\pm$ pairs surviving the topological and PID selections, a side-band subtraction procedure was used. This involved dividing the sample into $\langle p_T \rangle \in [5, 15] \text{ GeV}/c$ intervals and fitting the invariant-mass distributions of the D^0 candidates in each

interval with a Gaussian function for the signal and an exponential function for the background. The width (σ) and mean of the fitted Gaussian were used to define signal and side-band regions, with the two-dimensional distributions of θ and E_{Radiator} for D^0 -meson tagged jet candidates, $\langle \rho(\theta, E) \rangle$, obtained in each region. The signal region was defined to be within 2σ on either side of the Gaussian mean and contained most of the real D^0 mesons, with some contamination present from the combinatorial background. The side-band regions were defined to be from 4σ to 9σ away from the peak in either direction and were composed entirely of background D^0 -meson candidates. The combined $\langle \rho(\theta, E) \rangle$ distributions measured in the two side-band regions represent the structural form of the contribution of background candidates to the $\langle \rho(\theta, E) \rangle$ distribution measured in the signal region. In this way, the background component of the total $\langle \rho(\theta, E) \rangle$ measured in the signal region can be subtracted, using the following equation:

$$\langle \rho(\theta, E) \rangle = \sum_i \frac{1}{\omega} [\langle \rho(\theta, E) \rangle - \frac{A_S}{A_B} \langle \rho(\theta, E) \rangle_B], \quad (3)$$

where the subscripts S and B denote the signal and side-band regions of the invariant-mass distributions, respectively. The A_S and A_B variables are the areas under the background fit function in the signal and combined side-band regions, respectively, and were used to normalize the magnitude of the background in the side-band regions to that in the signal region. The D^0 -meson tagged jet selection efficiency (discussed in more detail in the next section) is denoted by ε , with the index i running over the $\langle p_T \rangle$ bins. As a result of this side-band

subtraction, the true D^0 -meson tagged jet $\langle \rho_{\theta, E} \rangle$ distributions are obtained, in the different intervals of $\langle p_T \rangle$.

D^0 -meson tagged jet reconstruction efficiency correction

The topological and PID selections used to identify the D^0 mesons, in the chosen jet kinematic interval, have a limited efficiency, which exhibits a strong p_T dependence. Therefore, before integrating the side-band subtracted $\langle \rho_{\theta, E} \rangle$ distributions across the measured $\langle p_T \rangle$ intervals, the $\langle \rho_{\theta, E} \rangle$ distributions were corrected for this efficiency. The efficiency, ε , was estimated from PYTHIA v.6 MC studies and varies strongly with $\langle p_T \rangle$, from approximately 0.01 at $\langle p_T \rangle = 2.5 \text{ GeV}/c$ to approximately 0.3 at $\langle p_T \rangle = 30 \text{ GeV}/c$ for prompt D^0 -meson tagged jets and from approximately 0.01 at $\langle p_T \rangle = 2.5 \text{ GeV}/c$ to approximately 0.2 at $\langle p_T \rangle = 30 \text{ GeV}/c$ for non-prompt D^0 -meson tagged jets. As the prompt and non-prompt D^0 -meson tagged jet reconstruction efficiencies were different, the final efficiency was obtained by combining the prompt and non-prompt D^0 -meson tagged jet reconstruction efficiencies, evaluated separately. These were combined with weights derived from simulations, corresponding to the admixture of prompt and non-prompt D^0 -meson tagged jets in the reconstructed sample. The fractions of this admixture were obtained in bins of $\langle p_T \rangle$ by calculating the prompt and non-prompt D^0 -meson tagged jet production cross sections with POWHEG combined with PYTHIA v.6 showering.

Evaluation of systematic uncertainties

Considered sources of systematic uncertainty in the measurement relate to the reconstruction and signal extraction of D^0 -meson candidates, with the

former contributing as the leading source. These uncertainties were estimated by varying the topological and PID selections, as well as the fitting and side-band subtraction configurations applied to the D^0 -meson candidate invariant mass distributions. Variations were chosen that tested the influence of selected analysis parameters as much as possible, while maintaining a reasonable significance in the signal extraction. For each of these categories, the root mean square of all deviations was taken as the final systematic uncertainty. Theoretical uncertainties in the prompt and non-prompt D^0 -meson tagged jet production cross sections from POWHEG were also considered in the calculation of the reconstruction efficiency, with the largest variation taken as the uncertainty. For each category, the final systematic uncertainty was symmetrized before adding up the uncertainties in quadrature across all categories to obtain the total systematic uncertainty of the D^0 -meson tagged jet measurement.

For the inclusive jet results, the minimum p_T requirement on the track with the highest transverse momentum within the leading prong of each splitting was varied. The magnitude of the variation was taken to be the resolution of the transverse momentum of a D^0 meson with $\langle p \rangle_{\{\{rm{T}\}\}}^{\{\{\rm{D}\}\}} = 2 \text{ GeV}/c$, which was found to be $0.06 \text{ GeV}/c$. Variations above and below the nominal selection value were made and the largest deviation was symmetrized. Systematic detector effects are dominated by the tracking efficiency and were shown in detector simulations to affect both the D^0 -meson tagged jet and inclusive jet samples equally, and they largely cancelled in the $R(\theta)$ ratio. Therefore, the systematic uncertainty of $R(\theta)$ because of detector effects was estimated directly on the ratio by randomly removing 15% of the reconstructed tracks, as given by the tracking efficiency of the ALICE detector, in the track samples used for clustering both the D^0 -meson tagged jets and inclusive jets. The ratio of the resulting $R(\theta)$ distribution to the case with no track removal was taken, to obtain the uncertainty, which was symmetrized.

The relative uncertainty of $R(\theta)$ resulting from the separate D^0 -meson tagged jet and inclusive jet uncertainties was calculated, with the resulting absolute uncertainty added in quadrature to the detector effects uncertainty to obtain the total systematic uncertainty of the $R(\theta)$ measurement. The magnitude of each of these sources of systematic uncertainty is shown in Table 1, for the

smallest-angle splittings corresponding to the interval $2 \leq \ln(1/\theta) < 3$, in which the uncertainties are largest.

Table 1 $R(\theta)$ systematic uncertainties

Data availability

All data shown in histograms and plots are publicly available on the HEPdata repository (<https://hepdata.net>).

Code availability

The source code utilized in this study is publicly available under the names AliPhysics and AliRoot. Further information can be provided by the authors upon reasonable request.

References

1. Particle Data Group Collaboration, Zyla, P. A. et al. Review of particle physics. *Prog. Theor. Exp. Phys.* **2020**, 083C01 (2020).
2. Buckley, A. et al. General-purpose event generators for LHC physics. *Phys. Rep.* **504**, 145–233 (2011).
3. Dokshitzer, Y. L., Khoze, V. A. & Troian, S. I. On specific QCD properties of heavy quark fragmentation ('dead cone'). *J. Phys. G* **17**, 1602–1604 (1991).
4. Frye, C., Larkoski, A. J., Thaler, J. & Zhou, K. Casimir meets Poisson: improved quark/gluon discrimination with counting observables. *J. High Energy Phys.* **9**, 083 (2017).
5. Dreyer, F. A., Salam, G. P. & Soyez, G. The Lund jet plane. *J. High Energy Phys.* **12**, 064 (2018).
6. S. Marzani, G. Soyez, and M. Spannowsky, *Looking Inside Jets: An Introduction to Jet Substructure and Boosted-Object Phenomenology*

Lecture Notes in Physics Vol. 958 (Springer, 2019).

7. DELPHI Collaboration, Abreu, P. et al. Hadronization properties of b quarks compared to light quarks in $\langle\{e\}^+\{e\}^-\rangle$ to $\overline{q}q$ from 183 GeV to 200 GeV. *Phys. Lett. B* **479**, 118–128 (2000). [Erratum: *Phys. Lett. B* **492**, 398–398 (2000)].
8. ATLAS Collaboration, Aad, G. et al. Measurement of jet shapes in top-quark pair events at $\sqrt{s} = 7$ TeV using the ATLAS detector. *Eur. Phys. J. C* **73**, 2676 (2013).
9. Llorente, J. & Cantero, J. Determination of the b -quark mass m_b from the angular screening effects in the ATLAS b -jet shape data. *Nucl. Phys. B* **889**, 401–418 (2014).
10. The SLD Collaboration, Abe, K. et al. Precise measurement of the b -quark fragmentation function in Z^0 boson decays. *Phys. Rev. Lett.* **84**, 4300–4304 (2000).
11. ALEPH Collaboration, Heister, A. et al. Study of the fragmentation of b quarks into B mesons at the Z peak. *Phys. Lett. B* **512**, 30–48 (2001).
12. OPAL Collaboration, Alexander, G. et al. A Study of b quark fragmentation into B^0 and B^+ mesons at LEP. *Phys. Lett. B* **364**, 93–106 (1995).
13. OPAL Collaboration, Akers, R. et al. A measurement of the production of D^\pm mesons on the Z^0 resonance. *Z. Phys. C* **67**, 27–44 (1995).
14. DELPHI Collaboration, Abreu, P. et al. A Measurement of B meson production and lifetime using $D\bar{L}^-$ events in Z_0 decays. *Z. Phys. C* **57**, 181–196 (1993).
15. CMS Collaboration, Sirunyan, A. M. et al. Measurement of jet substructure observables in $\langle\{\rm{t}\}\bar{\rm{t}}\rangle$ events from proton-proton collisions at $\sqrt{s}=13$ TeV. *Phys. Rev. D* **98**, 092014 (2018).

16. CMS Collaboration, Sirunyan, A. M. et al. Measurement of the splitting function in pp and Pb-Pb collisions at $\sqrt{s} = 5.02$ TeV. *Phys. Rev. Lett.* **120**, 142302 (2018).
17. ALICE Collaboration, Acharya, S. et al. Exploration of jet substructure using iterative declustering in pp and Pb–Pb collisions at LHC energies. *Phys. Lett. B* **802**, 135227 (2020).
18. STAR Collaboration, Abdallah, M. S. et al. Differential measurements of jet substructure and partonic energy loss in Au+Au collisions at $\sqrt{s} = 200$ GeV.
19. Larkoski, A. J., Marzani, S. & Thaler, J. Sudakov safety in perturbative QCD. *Phys. Rev. D* **91**, 111501 (2015).
20. ATLAS Collaboration, Aad, G. et al. Measurement of the Lund jet plane using charged particles in 13 TeV proton-proton collisions with the ATLAS detector. *Phys. Rev. Lett.* **124**, 222002 (2020).
21. Maltoni, F., Selvaggi, M. & Thaler, J. Exposing the dead cone effect with jet substructure techniques. *Phys. Rev. D* **94**, 054015 (2016).
22. Cunqueiro, L. & Ploskon, M. Searching for the dead cone effects with iterative declustering of heavy-flavor jets. *Phys. Rev. D* **99**, 074027 (2019).
23. ALICE Collaboration, Acharya, S. et al. Measurement of the production of charm jets tagged with D_0 mesons in pp collisions at $\sqrt{s} = 7$ TeV. *J. High Energy Phys.* **8**, 133 (2019).
24. Cacciari, M., Salam, G. P. & Soyez, G. The anti- k_t jet clustering algorithm. *J. High Energy Phys.* **04**, 063 (2008).
25. Cacciari, M., Salam, G. P. & Soyez, G. FastJet user manual. *Eur. Phys. J. C* **72**, 1896 (2012).

26. Dokshitzer, Y., Leder, G., Moretti, S. & Webber, B. Better jet clustering algorithms. *J. High Energy Phys.* **8**, 001 (1997).
27. Dokshitzer, Y., Khoze, V., Mueller, A. & Troyan, S. *Basics of Perturbative QCD* (Editions Frontieres, 1991).
28. Sjostrand, T., Mrenna, S. & Skands, P. Z. PYTHIA 6.4 Physics and Manual. *J. High Energy Phys.* **5**, 026 (2006).
29. Skands, P. Z. Tuning Monte Carlo generators: the Perugia tunes. *Phys. Rev. D* **82**, 074018 (2010).
30. Norrbin, E. & Sjostrand, T. QCD radiation off heavy particles. *Nucl. Phys. B* **603**, 297–342 (2001).
31. Brun, R., Bruyant, F., Maire, M., McPherson, A. C. & Zanarini, P. *GEANT 3: User's Guide Geant 3.10, Geant 3.11; Revised Version* (CERN, 1987); <https://cds.cern.ch/record/1119728>
32. Lifson, A., Salam, G. P. & Soyez, G. Calculating the primary Lund jet plane density. *J. High Energy Phys.* **10**, 170 (2020).
33. Sherpa Collaboration, Bothmann, E. et al. Event generation with Sherpa 2.2. *SciPost Phys.* **7**, 034 (2019).
34. Sjöstrand, T. et al. An introduction to PYTHIA 8.2. *Comput. Phys. Commun.* **191**, 159–177 (2015).
35. Frixione, S., Nason, P. & Ridolfi, G. A positive-weight next-to-leading-order Monte Carlo for heavy flavour hadroproduction. *J. High Energy Phys.* **9**, 126 (2007).
36. Dokshitzer, Y. L. & Kharzeev, D. Heavy quark colorimetry of QCD matter. *Phys. Lett. B* **519**, 199–206 (2001).
37. Armesto, N., Salgado, C. A. & Wiedemann, U. A. Medium induced gluon radiation off massive quarks fills the dead cone. *Phys. Rev. D* **69**, 114003 (2004).

38. Armesto, N., Dainese, A., Salgado, C. A. & Wiedemann, U. A. Testing the color charge and mass dependence of parton energy loss with heavy-to-light ratios at RHIC and CERN LHC. *Phys. Rev. D* **71**, 054027 (2005).
39. Bigi, I. I. Y., Dokshitzer, Y. L., Khoze, V. A., Kuhn, J. H. & Zerwas, P. M. Production and decay properties of ultraheavy quarks. *Phys. Lett. B* **181**, 157–163 (1986).
40. Cortiana, G. Top-quark mass measurements: review and perspectives. *Rev. Phys.* **1**, 60–76 (2016).
41. ALICE Collaboration, Abelev, B. et al. Performance of the ALICE experiment at the CERN LHC. *Int. J. Mod. Phys. A* **29**, 1430044 (2014).
42. ALICE Collaboration, Aamodt, K. et al. Alignment of the ALICE inner tracking system with cosmic-ray tracks. *J. Instrum.* **5**, P03003 (2010).
43. Alme, J. et al. The ALICE TPC, a large 3-dimensional tracking device with fast readout for ultra-high multiplicity events. *Nucl. Instrum. Meth. A* **622**, 316–367 (2010).
44. Akindinov, A. et al. Performance of the ALICE time-of-flight detector at the LHC. *Eur. Phys. J. Plus* **128**, 44 (2013).
45. ALICE Collaboration, Abbas, E. et al. Performance of the ALICE VZERO system. *J. Instrum.* **8**, P10016 (2013).
46. ALICE Collaboration, Abelev, B. et al. Performance of the ALICE experiment at the CERN LHC. *Int. J. Mod. Phys. A* **29**, 1430044 (2014).

Acknowledgements

The ALICE Collaboration is grateful to U. Wiedemann for his valuable suggestions and fruitful discussions. We are also grateful to D. Napoletano for providing the SHERPA configuration files and useful discussions. The ALICE Collaboration would like to thank all its engineers and technicians

for their invaluable contributions to the construction of the experiment and the CERN accelerator teams for the outstanding performance of the LHC complex. The ALICE Collaboration gratefully acknowledges the resources and support provided by all Grid centres and the Worldwide LHC Computing Grid (WLCG) collaboration. The ALICE Collaboration acknowledges the following funding agencies for their support in building and running the ALICE detector: A. I. Alikhanyan National Science Laboratory (Yerevan Physics Institute) Foundation (ANSL), State Committee of Science and World Federation of Scientists (WFS), Armenia; Austrian Academy of Sciences, Austrian Science Fund (FWF) (M 2467-N36) and Nationalstiftung für Forschung, Technologie und Entwicklung, Austria; Ministry of Communications and High Technologies, National Nuclear Research Center, Azerbaijan; Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Financiadora de Estudos e Projetos (Finep), Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) and Universidade Federal do Rio Grande do Sul (UFRGS), Brazil; Ministry of Education of China (MOEC), Ministry of Science & Technology of China (MSTC) and National Natural Science Foundation of China (NSFC), China; Ministry of Science and Education and Croatian Science Foundation, Croatia; Centro de Aplicaciones Tecnológicas y Desarrollo Nuclear (CEADEN), Cubaenergía, Cuba; Ministry of Education, Youth and Sports of the Czech Republic, Czech Republic; The Danish Council for Independent Research | Natural Sciences, the VILLUM FONDEN and Danish National Research Foundation (DNRF), Denmark; Helsinki Institute of Physics (HIP), Finland; Commissariat à l'Energie Atomique (CEA), Institut National de Physique Nucléaire et de Physique des Particules (IN2P3) and Centre National de la Recherche Scientifique (CNRS), France; Bundesministerium für Bildung und Forschung (BMBF) and GSI Helmholtzzentrum für Schwerionenforschung GmbH, Germany; General Secretariat for Research and Technology, Ministry of Education, Research and Religions, Greece; National Research, Development and Innovation Office, Hungary; Department of Atomic Energy Government of India (DAE), Department of Science and Technology, Government of India (DST), University Grants Commission, Government of India (UGC) and Council of Scientific and Industrial Research (CSIR), India; Indonesian Institute of Science, Indonesia; Istituto Nazionale di Fisica Nucleare (INFN), Italy; Institute for Innovative Science and Technology, Nagasaki Institute of

Applied Science (IIST), Japanese Ministry of Education, Culture, Sports, Science and Technology (MEXT) and Japan Society for the Promotion of Science (JSPS) KAKENHI, Japan; Consejo Nacional de Ciencia (CONACYT) y Tecnología, through Fondo de Cooperación Internacional en Ciencia y Tecnología (FONCICYT) and Dirección General de Asuntos del Personal Academico (DGAPA), Mexico; Nederlandse Organisatie voor Wetenschappelijk Onderzoek (NWO), the Netherlands; The Research Council of Norway, Norway; Commission on Science and Technology for Sustainable Development in the South (COMSATS), Pakistan; Pontificia Universidad Católica del Perú, Peru; Ministry of Education and Science, National Science Centre and WUT ID-UB, Poland; Korea Institute of Science and Technology Information and National Research Foundation of Korea (NRF), Republic of Korea; Ministry of Education and Scientific Research, Institute of Atomic Physics and Ministry of Research and Innovation and Institute of Atomic Physics, Romania; Joint Institute for Nuclear Research (JINR), Ministry of Education and Science of the Russian Federation, National Research Centre Kurchatov Institute, Russian Science Foundation and Russian Foundation for Basic Research, Russia; Ministry of Education, Science, Research and Sport of the Slovak Republic, Slovakia; National Research Foundation of South Africa, South Africa; Swedish Research Council (VR) and Knut & Alice Wallenberg Foundation (KAW), Sweden; European Organization for Nuclear Research, Switzerland; Suranaree University of Technology (SUT), National Science and Technology Development Agency (NSDTA) and Office of the Higher Education Commission under NRU project of Thailand, Thailand; Turkish Energy, Nuclear and Mineral Research Agency (TENMAK), Turkey; National Academy of Sciences of Ukraine, Ukraine; Science and Technology Facilities Council (STFC), United Kingdom; National Science Foundation of the United States of America (NSF) and United States Department of Energy, Office of Nuclear Physics (DOE NP), United States of America.

Author information

Author notes

1. G. Clai

Present address: Italian National Agency for New Technologies, Energy and Sustainable Economic Development, (ENEA), Bologna, Italy

2. M. Concas

Present address: Dipartimento DET del Politecnico di Torino, Turin, Italy

3. L. Malinina

Present address: D.V. Skobeltsyn Institute of Nuclear Physics, M.V. Lomonosov Moscow State University, Moscow, Russia

4. K. Redlich

Present address: Institute of Theoretical Physics, University of Wroclaw, Wroclaw, Poland

5. Deceased: J. Cleymans

6. Deceased: A. Grigoryan

7. Deceased: M. A. Mazzoni

8. Deceased: V. Samsonov

Affiliations

1. Variable Energy Cyclotron Centre, Homi Bhabha National Institute, Kolkata, India

S. Acharya, Z. Ahammed, P. P. Bhaduri, S. Chandra, S. Chattopadhyay, A. K. Dubey, P. Ghosh, D. Kumar, S. Muhuri, P. Pareek, S. K. Saha, J. Saini, N. Sarkar, V. K. Singh, V. Singhal & S. Thakur

2. Nuclear Physics Institute of the Czech Academy of Sciences, Řežu Prahy, Czech Republic

D. Adamova, J. Biel'cikova, A. Isakov, A. Kotliarov, F. Krizek, S. Kushpil & M. Šumbera

3. Institut fur Informatik, Fachbereich Informatik und Mathematik,
Johann-Wolfgang-Goethe Universitat Frankfurt, Frankfurt, Germany

A. Adler, T. Janson & U. Kebschull

4. Department of Physics, Division of Particle Physics, Lund University,
Lund, Sweden

J. Adolfsson, S. Basu, P. Christiansen, O. Matonoha, A. F.
Nassirpour, A. Ohlson, A. Oskarsson, T. Richert, O. V. Rueda & D.
Silvermyr

5. European Organization for Nuclear Research (CERN), Geneva,
Switzerland

G. Aglieri Rinella, A. Alkin, M. Angeletti, A. Augustinus, K. Barth, F.
Bellini, L. Betev, P. M. Bond, F. Carnesecchi, S. Chapeland, V.
Chibante Barroso, P. Chochula, M. Colocci, F. Costa, A. Di Mauro, R.
Divia, G. Eulisse, G. Fiorenza, U. Fuchs, C. Gargiulo, C. Grigoras, F.
Grosa, J. F. Grosse-Oetringhaus, J. A. Hasenbichler, H. Hillemanns, P.
Hristov, J. P. Iddon, G. M. Innocenti, N. Jacazio, J. M. Jowett, A.
Junque, A. Kalweit, M. Keil, J. Klein, A. Kluge, P. J. Konopka, V.
Kučera, S. Kundu, M. Lamanna, K. Lapidus, E. Laudi, L. Lautner, M.
Lettrich, M. Mager, P. Martinengo, A. L. Mazuecos, M. Mazzilli, A.
Morsch, T. Mrnjavac, L. Musa, R. N. Patra, O. Pinazza, M. Puccio, F.
Reidt, W. Riegler, D. Rohr, J. Schukraft, R. Shahoyan, G. Simonetti, M.
Suljic, A. Tauro, A. Telesca, S. Trogolo, P. Vande Vyvre, B. Volkel, B.
von Haller, A. Wegrzynek, S. C. Wenzel, C. Zampolli & N. Zardoshti

6. Dipartimento DISAT del Politecnico and Sezione INFN, Turin, Italy

M. Agnello, A. Balbino, S. Bufalino, F. Catalano, P. Fecchio & S.
Politano

7. INFN, Sezione di Bologna, Bologna, Italy

N. Agrawal, P. Antonioli, F. Cindolo, G. Clai, D. Hatzifotiadou, A. Margotti, R. Nania, F. Noferini, O. Pinazza, R. Preghenella & S. Tripathy

8. Department of Physics, Aligarh Muslim University, Aligarh, India

S. Ahmad, B. Ali, C. Anuj, M. D. Azmi, S. Khan, A. Khatun & M. Mohisin Khan

9. Korea Institute of Science and Technology Information, Daejeon, Republic of Korea

S. U. Ahn

10. Faculty of Science, P.J. Šafarik University, Košice, Slovakia

I. Ahuja, M. Bombara, A. Kravčáková, Z. Rescakova, M. Vala & J. Vrlakova

11. Indonesian Institute of Sciences, Jakarta, Indonesia

Z. Akbar & S. Sumowidagdo

12. NRC «Kurchatov» Institute – ITEP, Moscow, Russia

A. Akindinov, S. Kiselev, D. Mal'Kevich, K. Mikhaylov, S. Shirinkin, R. Sultanov, K. Voloshin & N. Zhigareva

13. Research Division and ExtreMe Matter Institute EMMI, GSI Helmholtzzentrum für Schwerionenforschung GmbH, Darmstadt, Germany

M. Al-Turany, R. Averbeck, X. Bai, M. B. Blidaru, P. Braun-Munzinger, A. Caliva, M. R. Ciupek, A. Dubla, P. Foka, C. Garabatos, P. Gasik, P. Giubellino, R. Gross, M. K. Habib, A. Harlenderova, S. Hornung, M. Ivanov, J. M. Jowett, T. Kollegger, L. Kreis, C. Lippmann, P. Malzacher, A. Marin, S. Masciocchi, D. Miskowiec, C. Schmidt, K. Schwarz, K. Schweda, I. Selyuzhenkov, J. Wilkinson & J. Zhu

14. Fudan University, Shanghai, China

S. N. Alam, Y. G. Ma, Q. Shou, C. Wang, D. Wang, W. Wu & S. Zhang

15. National Research Centre Kurchatov Institute, Moscow, Russia

D. Aleksandrov, D. Blau, S. Fokin, M. Ippolitov, A. Kazantsev, V. Manko, S. Nikolaev, S. Nikulin, A. Nyanin, D. Peresunko, E. Ryabinkin, Y. Sibiriak, A. Vasiliev & A. Vinogradov

16. INFN, Sezione di Torino, Turin, Italy

B. Alessandro, R. Arnaldi, M. Concas, N. De Marco, A. Feliciello, P. Giubellino, F. Grossa, L. Micheletti, C. Oppedisano, F. Prino, E. Scomparin & S. Vallero

17. Central China Normal University, Wuhan, China

H. M. Alfanda, M. N. Anaam, M. Cai, P. Cui, W. Deng, Y. Ding, F. Fan, M. Hamid, S. Kar, A. M. Khan, X. L. Li, Y. Mao, H. Pei, X. Peng, S. Tang, R. Xu, Z. Yin, X. Zhang, D. Zhou, J. Zhu & Y. Zhu

18. Instituto de Fisica, Universidad Nacional Autonoma de Mexico, Mexico City, Mexico

R. Alfaro Molina, L. H. Gonzalez-Trueba, V. Grabski & A. Menchaca-Rocha

19. COMSATS University Islamabad, Islamabad, Pakistan

Y. Ali, J. B. Butt, H. Ilyas, M. U. Naru, U. Tabassam & A. Zaman

20. Dipartimento di Fisica e Astronomia dell'Universita and Sezione INFN, Bologna, Italy

A. Alici, S. Arcelli, F. Bellini, F. Carnesecchi, L. Cifarelli, M. Colocci, F. Ercolelli, M. Giacalone, N. Rubini, G. Scioli & A. Zichichi

21. University of Houston, Houston, TX, USA

N. Alizadehvandchali, R. Bellwied, F. Flor, A. Hutson, A. G. Knospe, J. L. Martinez, A. S. Menon, S. P. Pathak, L. Pinsky, R. E. Quishpe, O. Sheibani, J. Song, C. Terrevoli & A. R. Timmins

22. Department of Physics and Technology, University of Bergen, Bergen, Norway

J. Alme, L. Altenkamper, O. Djupsland, V. N. Eikeland, M. R. Ersdal, F. M. Fionda, O. S. Groettvik, I. M. Lofnes, J. Nystrand, A. Rehman, D. Rohrich, J. Saetre, G. J. Tambave, K. Ullaland, B. Wagner, S. Yang & S. Yuan

23. Institut fur Kernphysik, Johann Wolfgang Goethe-Universitat Frankfurt, Frankfurt, Germany

T. Alt, H. Appelshauser, R. Bailhache, E. Bartsch, C. Blume, L. Bratrud, H. Buesching, B. Donigus, F. Eisenhut, G. Halimoglu, E. Hellbar, P. Huhn, J. Jung, M. Jung, S. Kirsch, M. Kleiner, J. Konig, M. Kruger, M. Marquard, A. F. Mechler, R. H. Munzer, F. Pliquet, P. Reichelt, R. Renfordt, T. S. Rogoschinski, H. S. Scheid, N. V. Schmidt, A. Toia & J. Wiechula

24. St. Petersburg State University, St. Petersburg, Russia

I. Altsybeev, S. Belokurova, A. Erokhin, G. Feofilov, V. Kovalenko, T. Lazareva, D. Nesterov, V. Vechernin, A. Zarochentsev & V. Zherebchevskii

25. Horia Hulubei National Institute of Physics and Nuclear Engineering, Bucharest, Romania

C. Andrei, A. Bercuci, A. Lindner, M. Petrovici, A. Pop, C. Schiaua & M. G. Tarzila

26. Nikhef, National Institute for Subatomic Physics, Amsterdam, the Netherlands

D. Andreou, D. Caffarri, P. Christakoglou, A. Dubla, B. Hohlweger, Z. Khabanova, P. G. Kuijer, S. Qiu, I. Ravasenga, C. P. Stylianidis, M. van

Leeuwen & R. J. G. V. Weelden

27. Institut fur Kernphysik, Westfälische Wilhelms-Universitat Munster,
Munster, Germany

A. Andronic, A. D. C. Bell Hechavarria, T. M. Eder, K. Garner, F. Herrmann, J. Honermann, L. A. Husova, F. Jonas, C. Klein-Bosing, J. R. Luhder, D. A. Moreira De Godoy, D. Muhlheim, J. P. Wessels & G. A. Willems

28. Physikalisches Institut, Ruprecht-Karls-Universitat Heidelberg,
Heidelberg, Germany

V. Anguelov, M. Arslandok, A. Berdnikova, L. Bergmann, M. C. Danisch, V. J. G. Feuillard, P. Glassel, J. Kim, M. Kim, M. K. Kohler, M. Kroesen, J. A. Lopez, N. A. Martin, A. Nath, Y. Pachmayer, K. Reygers, R. Schicker, A. Schmah, M. O. Schmidt, G. Skorodumovs, J. Stachel, S. F. Stiefelmaier, M. A. Volkl, B. Windelband & A. Yuncu

29. INFN, Sezione di Padova, Padova, Italy

F. Antinori, A. Dainese, A. Rossi & R. Turrisi

30. Lawrence Berkeley National Laboratory, Berkeley, CA, USA

N. Apadula, A. Collu, R. Cruz-Torres, L. Greiner, B. Jacak, P. M. Jacobs, S. Klein, Y. S. Lai, J. D. Mulligan, M. Płoskoń & J. Porter

31. SUBATECH, IMT Atlantique, Universite de Nantes, CNRS-IN2P3,
Nantes, France

L. Aphecetche, G. Batigne, O. Bugnon, F. P. A. Damas, B. Erazmus, K. Garg, M. Germain, M. Guilbaud, G. Martinez Garcia, R. P. Pezzi, P. Pillot, R. Sadek, A. Shabetai, D. Stocco & M. Tarhini

32. Department of Physics, University of Oslo, Oslo, Norway

I. C. Arsene, O. Dordic, Q. W. Malik, A. Neagu, M. Richter, K. Roed, T. B. Skaali, I. Storehaug, T. S. Tveter & J. Wikne

33. Yale University, New Haven, CT, USA

M. Arslanbekov, C. Beattie, H. Bossi, H. Caines, J. W. Harris, L. B. Havener, M. H. Oliver, M. H. P. Sas & N. Smirnov

34. Laboratoire de Physique des 2 Infinis, Irene Joliot-Curie, Orsay, France

S. Aziz, Z. Conesa del Valle, B. Espagnon, C. Hadjidakis, L. Massacrier, C. Suire & T. Tork

35. INFN, Sezione di Catania, Catania, Italy

A. Badala, G. Mandaglio, A. Rosano & A. Trifiro

36. Gangneung-Wonju National University, Gangneung, Republic of Korea

Y. W. Baek & J. S. Kim

37. University of Science and Technology of China, Hefei, China

X. Bai, Z. Tang, Y. Wu & Y. Zhang

38. Indian Institute of Technology Indore, Indore, India

Y. Bailung, S. Deb, N. Mallick, R. Rath, S. P. Rode, A. Roy, R. Sahoo, D. Sahu & R. Singh

39. Physics Department, University of Jammu, Jammu, India

R. Bala, A. Bhasin, A. Gupta, R. Gupta, A. Jalotra, N. M. Malik, R. N. Patra, S. Sambyal, M. Sharma, S. Sharma, U. Sharma, R. Singh & V. Sumberia

40. Department de Physique Nucleaire (DPhN), Universite Paris-Saclay
Centre d'Etudes de Saclay (CEA), IRFU, Saclay, France

A. Baldisseri, H. Borel, R. Caron, J. Castillo Castellanos, M. L. Coquet, F. P. A. Damas, A. Ferrero, A. M. C. Glaenzer, S. Panebianco, H. Pereira Da Costa, S. Perrin, A. Rakotozafindrabe & M. Winn

41. AGH University of Science and Technology, Cracow, Poland

B. Balis, M. Gorgon, A. Horzyk, M. Jablonski, J. P. Kitowski, P. J. Konopka, S. D. Koryciak & P. G. Russek

42. Helmholtz-Institut fur Strahlen- und Kernphysik, Rheinische Friedrich-Wilhelms-Universitat Bonn, Bonn, Germany

M. Ball, P. Hauer, B. Ketzer & T. Mahmoud

43. Bose Institute, Department of Physics and Centre for Astroparticle Physics and Space Science (CAPSS), Kolkata, India

D. Banerjee, M. A. Bhat, S. Biswas, P. Das, S. Das, S. K. Ghosh, A. Modak & S. K. Prasad

44. Dipartimento di Fisica e Astronomia dell'Universita and Sezione INFN, Catania, Italy

R. Barbera, P. La Rocca, C. Pinto & F. Riggi

45. Physik Department, Technische Universitat Munchen, Munich, Germany

L. Barioglio, A. Bilandzic, L. Fabbietti, S. T. Heckel, B. Hohlweger, T. Klemenz, S. A. Konigstorfer, L. Lautner, M. Lettrich, A. M. Mathis, D. L. Mihaylov, C. Mordasini, V. M. Sarti, L. Šerkšnytė, B. Singh, S. F. Taghavi, O. Vazquez Doce & I. Vorobyev

46. Dipartimento di Fisica dell'Universita and Sezione INFN, Turin, Italy

L. Barioglio, S. Beole, L. Bianchi, E. Botta, L. De Cilladi, S. Delsanto, A. Ferretti, M. Gagliardi, M. Masera, F. Mazzaschi, L. Micheletti, L. Quaglia, L. Terlizzi & E. Vercellin

47. Department of Physics, School of Science, National and Kapodistrian University of Athens,, Athens, Greece

M. Barlou, P. Ganoti & M. Vasileiou

48. Wigner Research Centre for Physics, Budapest, Hungary

G. G. Barnafoldi, G. Biro, L. Boldizsar, D. Colella, E. Frajna, L. Gyulai, G. Hamar, P. Levai, A. N. Mishra, D. Varga, Z. Varga, M. Varga-Kofarago & R. Vertesi

49. Nuclear Physics Group, STFC Daresbury Laboratory, Daresbury, UK

L. S. Barnby & R. C. Lemmon

50. Universite Clermont Auvergne, CNRS/IN2P3, LPC, Clermont-Ferrand, France

V. Barret, N. Bastid, P. Crochet, P. Dupieux, X. Lopez, F. Manso, S. Porteboeuf-Houssais, G. Taillepied & S. Tang

51. University of Liverpool, Liverpool, UK

C. Bartels, M. D. Buckland, M. Chartier, C. Hills, J. P. Iddon, J. Liu & J. Norman

52. Dipartimento di Fisica e Astronomia dell'Universita and Sezione INFN, Padova, Italy

F. Baruffaldi, M. Cai, M. Faggin, P. Giubilato, M. Lunardon, F. Soramel & S. Trogolo

53. Joint Institute for Nuclear Research (JINR), Dubna, Russia

B. Batyunya, C. Ceballos Sanchez, S. Grigoryan, A. Kondratyev, L. Malinina, K. Mikhaylov, P. Nomokonov, V. Pozdniakov, E. Rogochaya, B. Rumyantsev & A. Vodopyanov

54. Indian Institute of Technology Bombay (IIT), Mumbai, India

D. Bauri, P. Chakraborty, S. Dash, B. Naik, B. K. Nandi, S. Padhan, S. Roy, B. Sahoo, P. Sahoo & T. Tripathy

55. Seccion Fisica, Departamento de Ciencias, Pontificia Universidad Catolica del Peru, Lima, Peru

J. L. Bazo Alba, E. Calvo Villar, A. M. Gago & C. Soncco

56. Niels Bohr Institute, University of Copenhagen, Copenhagen, Denmark

I. G. Bearden, C. H. Christensen, J. J. Gaardhoje, K. Gulbrandsen, Z. Moravcova, B. S. Nielsen, T. Richert, V. Vislavicius & Y. Zhou

57. Universite de Strasbourg, CNRS, IPHC UMR 7178, Strasbourg, France

I. Belikov, A. Gal, B. Hippolyte, C. Kuhn, A. Maire, F. Rami, R. Schotter, Y. Schutz & S. Senyukov

58. NRNU Moscow Engineering Physics Institute, Moscow, Russia

V. Belyaev, A. Bogdanov, A. Bolozdynya, V. Grigoriev, V. Kaplin, N. Kondratyeva, V. Loginov, V. A. Okorokov, V. Samsonov & I. Selyuzhenkov

59. Instituto de Ciencias Nucleares, Universidad Nacional Autonoma de Mexico, Mexico City, Mexico

G. Bencedi, E. Cuautle, L. Nellen, A. Ortiz Velasquez, G. Pai' c, E. D. Rosas & S. Tripathy

60. Petersburg Nuclear Physics Institute, Gatchina, Russia

Y. Berdnikov, V. Ivanov, A. Khanzadeev, E. Kryshen, M. Malaev, V. Nikulin, A. Riabov, V. Riabov, Y. Ryabov, V. Samsonov & M. Zhalov

61. Institute of Space Science (ISS), Bucharest, Romania

M. G. Besoiu, A. Danu, A. Dobrin, C. Ristea, A. Sevcenco & I. Stan

62. Department of Physics, Gauhati University, Guwahati, India

B. Bhattacharjee & P. Sarma

63. Dipartimento di Fisica dell'Universita and Sezione INFN, Cagliari, Italy

P. Bhattacharya, S. Boi, E. A. R. Casula, A. Chauvin, A. De Falco, A. Mulliri, B. Paul & G. L. Usai

64. INFN, Laboratori Nazionali di Frascati, Frascati, Italy

N. Bianchi, A. Fantoni, P. Gianotti, P. Larionov, V. Muccifora, S. Pisano, F. Ronchetti & M. Toppi

65. Faculty of Nuclear Sciences and Physical Engineering, Czech Technical University in Prague, Prague, Czech Republic

J. Biel'cik, M. Broz, J. G. Contreras, T. Herman, D. Horak, A. Isakov, K. Krizkova Gajdosova, R. Lavicka, V. Petráček, K. Šafařík, S. R. Torres & B. A. Trzeciak

66. The Henryk Niewodniczanski Institute of Nuclear Physics, Polish Academy of Sciences, Cracow, Poland

J. Biernat, S. A. Bysiak, J. Figiel, L. Gorlich, A. Khuntia, M. Kowalski, A. Matyja, C. Mayer, J. Otwinowski, A. Rybicki, H. Sharma & I. Sputowska

67. The University of Texas at Austin, Austin, TX, USA

J. T. Blair, A. N. Flores, E. F. Gauger, R. Hannigan, C. Markert, J. Schambach, D. Thomas & J. R. Wright

68. Dipartimento di Fisica e Nucleare e Teorica, Universita di Pavia, Pavia, Italy

G. Boca, S. Costanza, A. Rotondi & N. Valle

69. INFN, Sezione di Pavia, Pavia, Italy

G. Boca, G. Bonomi, S. Costanza, R. Lea, D. Pagano, J. Pazzini, A. Rotondi, M. Urioni, N. Valle & N. Zurlo

70. Oak Ridge National Laboratory, Oak Ridge, TN, USA

F. Bock, T. M. Cormier, L. Cunqueiro, R. J. Ehlers, M. Fasel, H. Hassan, F. Jonas, C. Loizides, M. G. Poghosyan, K. F. Read, J. Schambach & N. V. Schmidt

71. Inha University, Incheon, Republic of Korea

J. Bok, S. Cho, J. Do, B. Kim, M. J. Kweon, J. Y. Kwon, J. Park, J. J. Seo & J. H. Yoon

72. Universita di Brescia, Brescia, Italy

G. Bonomi, R. Lea, D. Pagano, J. Pazzini, M. Urioni & N. Zurlo

73. Moscow Institute for Physics and Technology, Moscow, Russia

A. Borissov

74. Universidade de Sao Paulo (USP), Sao Paulo, Brazil

M. Bregant, F. D. M. Canedo, H. F. Degenhardt, C. Jahnke, P. F. T. Matuoka, M. G. Munhoz, T. F. Silva & A. A. P. Suaide

75. Politecnico di Bari and Sezione INFN, Bari, Italy

G. E. Bruno, D. Colella & G. Fiorenza

76. Dipartimento Interateneo di Fisica ‘M. Merlin’ and Sezione INFN, Bari, Italy

G. E. Bruno, D. Colella, D. Di Bari, A. G. T. Ramos, G. Trombetta & G. Volpe

77. Russian Federal Nuclear Center (VNIIEF), Sarov, Russia

D. Budnikov, S. Filchagin, R. Ilkaev, A. Kuryakin, A. Tumkin & N. Zaviyalov

78. Stefan Meyer Institut fur Subatomare Physik (SMI), Vienna, Austria

P. Buhler, E. Meninno & M. Weber

79. iThemba LABS, National Research Foundation, Somerset West, South Africa

Z. Buthelezi, S. Foertsch, S. Mhlanga & T. J. Shaba

80. University of the Witwatersrand, Johannesburg, South Africa

Z. Buthelezi, J. E. Mdhluli, B. Naik & Z. Vilakazi

81. Universidad Autonoma de Sinaloa, Culiacan, Mexico

J. M. M. Camacho, C. D. Galvan & I. Leon Monzon

82. High Energy Physics Group, Universidad Autonoma de Puebla, Puebla, Mexico

R. S. Camacho, T. G. Chavez, A. Fernandez Tellez, J. R. A. Garcia, S. P. Guzman, E. G. Hernandez, M. I. Martinez, L. A. P. Moreno, S. A. R. Ramirez, M. Rodriguez Cahuantzi, P. F. Rojas, G. Tejeda Munoz, A. Vargas & S. Vergara Limon

83. Dipartimento di Fisica dell'Universita and Sezione INFN, Trieste, Italy

P. Camerini, G. Contin, C. De Martin, R. Lea, G. V. Margagliotti, R. Rui & V. Zaccolo

84. Saha Institute of Nuclear Physics, Homi Bhabha National Institute, Kolkata, India

S. Chattopadhyay, I. Das, H. Hushnud, M. S. Islam, P. Roy, W. Shaikh & T. Sinha

85. Universite de Lyon, CNRS/IN2P3, Institut de Physique des 2 Infinis de Lyon, Lyon, France
C. Cheshkov, B. Cheynis, Y. Ding, L. C. Migliorin, R. Tieulent & A. Uras
86. Universidade Estadual de Campinas (UNICAMP), Campinas, Brazil
D. D. Chinellato, G. G. Guardiano, C. Jahnke & J. Takahashi
87. University of Tsukuba, Tsukuba, Japan
T. Chujo, M. Inaba, T. Kumaoka, Y. Miake, N. Novitzky & S. Sakai
88. INFN, Sezione di Cagliari, Cagliari, Italy
C. Cicalo, F. M. Fionda, A. Masoni & S. Siddhanta
89. University of Cape Town, Cape Town, South Africa
J. Cleymans, T. Dietel, S. Mhlanga & S. Murray
90. INFN, Sezione di Bari, Bari, Italy
F. Colamaria, D. Colella, G. de Cataldo, D. Elia, V. Manzari, A. Mastroserio, E. Nappi, A. Palasciano & G. Vino
91. School of Physics and Astronomy, University of Birmingham, Birmingham, UK
J. S. Colburn, D. Evans, O. Jevons, P. G. Jones, A. Jusko, M. Krivda, J. Kvapil, R. Lietava, S. Ragoni & O. Villalobos Baillie
92. Laboratoire de Physique Subatomique et de Cosmologie, Universite Grenoble-Alpes, CNRS-IN2P3, Grenoble, France
G. Conesa Balbastre, J. Faivre, C. Furget, R. Guernane & T. A. Rancien
93. Dipartimento di Scienze e Innovazione Tecnologica dell'Universita del Piemonte Orientale and INFN Sezione di Torino, Alessandria, Italy

P. Cortese, L. Ramello & M. Sitta

94. Universidade Federal do ABC, Santo Andre, Brazil

M. R. Cosentino

95. National Institute of Science Education and Research, Homi Bhabha
National Institute, Jatni, India

P. Das, S. De, M. Dukhishyam, M. R. Haque, S. Kundu, B. Mohanty, S.
Saha & R. Singh

96. Dipartimento di Fisica ‘E.R. Caianiello’ dell’Universita and Gruppo
Collegato INFN, Salerno, Italy

A. De Caro, D. De Gruttola, S. De Pasquale, L. Dello Stritto, N.
Funicello, E. Meninno & T. Virgili

97. Frankfurt Institute for Advanced Studies, Johann Wolfgang Goethe-
Universitat Frankfurt, Frankfurt, Germany

J. de Cuveland, S. Gorbunov, D. Hutter, I. Kisel, M. Krzewicki, S. L.
La Pointe, J. Lehrbach, V. Lindenstruth, G. Neskovic & A. R.
Redelbach

98. Warsaw University of Technology, Warsaw, Poland

K. R. Deja, L. K. Graczykowski, M. R. Haque, M. J. Jakubowska, M.
A. Janik, P. Karczmarczyk, A. Kisiel, G. Kornakov, J. W. Myrcha, P.
Nowakowski, J. Oleniacz, P. S. Rokita, W. Rzesz, T. P. Trzcinski & H.
Zbroszczyk

99. Department of Physics, University of California, Berkeley, CA, USA

P. Dhankher, D. U. Dixit, E. D. Lesser, A. Liu & F. Torales-Acosta

100. Centro de Aplicaciones Tecnologicas y Desarrollo Nuclear (CEADEN),
Havana, Cuba

R. A. Diaz, E. Lopez Torres & G. M. Perez

101. Institute for Nuclear Research, Academy of Sciences, Moscow, Russia

U. Dmitrieva, D. Finogeev, A. Furs, O. Karavichev, T. Karavicheva, E. Karpechev, A. Kurepin, A. B. Kurepin, A. Maevskaya, Y. Melikyan, I. Morozov, I. Pshenichnov, D. Serebryakov, A. Shabanov, A. Tikhonov, N. Topilskaya & N. Vozniuk

102. Physics Department, Panjab University, Chandigarh, India

S. Dudi, L. Kumar, N. Kumar, A. Sharma & N. Sharma

103. Comenius University Bratislava, Faculty of Mathematics, Physics and Informatics, Bratislava, Slovakia

N. Dzalaiova, M. Meres, M. Pikna, B. Sitar, A. Szabo & I. Szarka

104. Physics Department, Faculty of Science, University of Zagreb, Zagreb, Croatia

F. Erhardt, M. Jercic, D. Karatovic, M. Planinic & N. Poljak

105. NRC Kurchatov Institute IHEP, Protvino, Russia

S. Evdokimov, V. Izucheev, Y. Kharlov, E. Kondratyuk, B. Polichtchouk, R. Rogalev, S. Sadovsky & A. Shangaraev

106. INFN, Sezione di Trieste, Trieste, Italy

E. Fragiacomo, G. Luparello & S. Piano

107. Chicago State University, Chicago, IL, USA

E. Garcia-Solis & A. Harton

108. National Nuclear Research Center, Baku, Azerbaijan

A. Garibli & A. Rustamov

109. University of Kansas, Lawrence, KS, USA
A. Gautam
110. Instituto de Fisica, Universidade Federal do Rio Grande do Sul (UFRGS), Porto Alegre, Brazil
M. B. Gay Ducati, L. G. Pereira & R. P. Pezzi
111. Nagasaki Institute of Applied Science, Nagasaki, Japan
D. J. Q. Goh, H. Hamagaki, M. Ogino & K. Oyama
112. Wayne State University, Detroit, MI, USA
V. Gonzalez, J. Pan, C. A. Pruneau, D. Sarkar & S. A. Voloshin
113. Faculty of Electrical Engineering, Mechanical Engineering and Naval Architecture, University of Split, Split, Croatia
S. Gotovac, P. Loncar, E. Mudnic & L. Vickovic
114. Institute for Gravitational and Subatomic Physics (GRASP), Utrecht University/Nikhef, Utrecht, the Netherlands
A. Grelli, B. Hofman, S. Jaelani, D. L. D. Keijdener, A. P. Mohanty, T. Peitzmann, R. J. M. Snellings, N. van der Kolk, L. V. R. van Doremalen, L. Vermunt, M. Verweij, H. Yokoyama & H. J. C. Zanolli
115. A.I. Alikhanyan National Science Laboratory (Yerevan Physics Institute) Foundation, Yerevan, Armenia
A. Grigoryan & S. Grigoryan
116. University of Tokyo, Tokyo, Japan
T. Gunji, S. Hayashi, H. Murakami, Y. Sekiguchi & D. Sekihata
117. Faculty of Engineering and Science, Western Norway University of Applied Sciences, Bergen, Norway

H. Helstrup, K. F. Hetland, B. Kileng, S. V. Nesbo, H. A. Remeness & M. M. Storetvedt

118. Centro de Investigacion y de Estudios Avanzados (CINVESTAV), Mexico City and Merida, Mexico

G. Herrera Corral

119. Yonsei University, Seoul, Republic of Korea

G. H. Hong, D. Kim, J. Kim, T. Kim & Y. Kwon

120. Creighton University, Omaha, NE, USA

R. Hosokawa & J. E. Seger

121. University of Tennessee, Knoxville, TN, USA

C. Hughes, C. Nattrass, A. C. Oliveira Da Silva, K. F. Read, A. R. Schmier, S. Sorensen, P. J. Steffanic & W. E. Witt

122. Ohio State University, Columbus, OH, USA

T. J. Humanic

123. Technical University of Košice, Košice, Slovakia

S. Jadlovska, J. Jadlovsky, L. Koska, M. Tkacik & D. Voscek

124. Suranaree University of Technology, Nakhon Ratchasima, Thailand

J. Kaewjai, C. Kobdaj, A. Lakrathok & A. Songmoolnak

125. Institute of Experimental Physics, Slovak Academy of Sciences, Košice, Slovakia

P. Kalinak, I. Kralik, M. Krivda & J. Musinsky

126. KTO Karatay University, Konya, Turkey

A. Karasu Uysal & S. Yalcin

127. Hochschule Worms, Zentrum fur Technologietransfer und Telekommunikation (ZTT), Worms, Germany

R. Keidel

128. Department of Physics, Pusan National University, Pusan, Republic of Korea

B. Kim, C. Kim, B. Lim, S. H. Lim & I.-K. Yoo

129. University of Jyvaskyla, Jyvaskyla, Finland

D. J. Kim, A. Onnerstad, J. E. Parkkila, H. Rytkonen & W. H. Trzaska

130. Jeonbuk National University, Jeonju, Republic of Korea

E. J. Kim & J. Kim

131. Department of Physics, Sejong University, Seoul, Republic of Korea

S. Kim

132. California Polytechnic State University, San Luis Obispo, CA, USA

J. L. Klay

133. National Centre for Nuclear Research, Warsaw, Poland

O. Kovalenko, P. Kurashvili, R. Nair, K. Redlich, T. Siemianczuk & G. Wilk

134. University of South-Eastern Norway, Tonsberg, Norway

R. Langoy & J. Lien

135. China Institute of Atomic Energy, Beijing, China

X. Li & Y. Zhi

136. Dipartimento di Scienze MIFT, Universita di Messina, Messina, Italy

G. Mandaglio, A. Rosano & A. Trifiro

137. Institute of Physics of the Czech Academy of Sciences, Prague, Czech Republic

J. Mareš & P. Zavada

138. Universita degli Studi di Foggia, Foggia, Italy

A. Mastroserio

139. INFN, Sezione di Roma, Rome, Italy

M. A. Mazzoni

140. Dipartimento di Fisica dell'Universita 'La Sapienza' and Sezione INFN, Rome, Italy

F. Meddi

141. Chungbuk National University, Cheongju, Republic of Korea

S. Noh

142. Hiroshima University, Hiroshima, Japan

T. Osako, K. Shigaki, T. Sugitate, Y. Yamaguchi, K. Yamakawa & S. Yano

143. Budker Institute for Nuclear Physics, Novosibirsk, Russia

Y. Pestov

144. Physics Department, University of Rajasthan, Jaipur, India

R. Raniwala & S. Raniwala

145. Helsinki Institute of Physics (HIP), Helsinki, Finland

S. S. Rasanen, O. A. M. Saarimaki & M. Slupecki

146. Institute of Physics, Homi Bhabha National Institute, Bhubaneswar, India

S. Sahoo, P. K. Sahu & S. Swain

147. Physikalisches Institut, Eberhard-Karls-Universitat Tubingen, Tubingen, Germany

H. R. Schmidt & M. Schmidt

148. Nara Women's University (NWU), Nara, Japan

M. Shimomura

149. Bogolyubov Institute for Theoretical Physics, National Academy of Sciences of Ukraine, Kiev, Ukraine

G. Tersimonov, V. Trubnikov & G. Zinovjev

Consortia

ALICE Collaboration

- S. Acharya
- , D. Adamova
- , A. Adler
- , J. Adolfsson
- , G. Aglieri Rinella
- , M. Agnello
- , N. Agrawal
- , Z. Ahammed
- , S. Ahmad
- , S. U. Ahn
- , I. Ahuja
- , Z. Akbar
- , A. Akindinov

- , M. Al-Turany
- , S. N. Alam
- , D. Aleksandrov
- , B. Alessandro
- , H. M. Alfanda
- , R. Alfaro Molina
- , B. Ali
- , Y. Ali
- , A. Alici
- , N. Alizadehvandchali
- , A. Alkin
- , J. Alme
- , T. Alt
- , L. Altenkamper
- , I. Altsybeev
- , M. N. Anaam
- , C. Andrei
- , D. Andreou
- , A. Andronic
- , M. Angeletti
- , V. Anguelov
- , F. Antinori
- , P. Antonioli
- , C. Anuj
- , N. Apadula
- , L. Aphecetche
- , H. Appelshauser
- , S. Arcelli
- , R. Arnaldi
- , I. C. Arsene
- , M. Arslandok
- , A. Augustinus
- , R. Averbeck
- , S. Aziz
- , M. D. Azmi
- , A. Badala
- , Y. W. Baek

- , X. Bai
- , R. Bailhache
- , Y. Bailung
- , R. Bala
- , A. Balbino
- , A. Baldisseri
- , B. Balis
- , M. Ball
- , D. Banerjee
- , R. Barbera
- , L. Barioglio
- , M. Barlou
- , G. G. Barnafoldi
- , L. S. Barnby
- , V. Barret
- , C. Bartels
- , K. Barth
- , E. Bartsch
- , F. Baruffaldi
- , N. Bastid
- , S. Basu
- , G. Batigne
- , B. Batyanya
- , D. Bauri
- , J. L. Bazo Alba
- , I. G. Bearden
- , C. Beattie
- , I. Belikov
- , A. D. C. Bell Hechavarria
- , F. Bellini
- , R. Bellwied
- , S. Belokurova
- , V. Belyaev
- , G. Bencedi
- , S. Beole
- , A. Bercuci
- , Y. Berdnikov

- , A. Berdnikova
- , L. Bergmann
- , M. G. Besoiu
- , L. Betev
- , P. P. Bhaduri
- , A. Bhasin
- , M. A. Bhat
- , B. Bhattacharjee
- , P. Bhattacharya
- , L. Bianchi
- , N. Bianchi
- , J. Biel'cik
- , J. Biel'cikova
- , J. Biernat
- , A. Bilandzic
- , G. Biro
- , S. Biswas
- , J. T. Blair
- , D. Blau
- , M. B. Blidaru
- , C. Blume
- , G. Boca
- , F. Bock
- , A. Bogdanov
- , S. Boi
- , J. Bok
- , L. Boldizsar
- , A. Bolozdynya
- , M. Bombara
- , P. M. Bond
- , G. Bonomi
- , H. Borel
- , A. Borissov
- , H. Bossi
- , E. Botta
- , L. Bratrud
- , P. Braun-Munzinger

- , M. Bregant
- , M. Broz
- , G. E. Bruno
- , M. D. Buckland
- , D. Budnikov
- , H. Buesching
- , S. Bufalino
- , O. Bugnon
- , P. Buhler
- , Z. Buthelezi
- , J. B. Butt
- , S. A. Bysiak
- , D. Caffarri
- , M. Cai
- , H. Caines
- , A. Caliva
- , E. Calvo Villar
- , J. M. M. Camacho
- , R. S. Camacho
- , P. Camerini
- , F. D. M. Canedo
- , F. Carnesecchi
- , R. Caron
- , J. Castillo Castellanos
- , E. A. R. Casula
- , F. Catalano
- , C. Ceballos Sanchez
- , P. Chakraborty
- , S. Chandra
- , S. Chapelard
- , M. Chartier
- , S. Chattopadhyay
- , S. Chattopadhyay
- , A. Chauvin
- , T. G. Chavez
- , C. Cheshkov
- , B. Cheynis

- , V. Chibante Barroso
- , D. D. Chinellato
- , S. Cho
- , P. Chochula
- , P. Christakoglou
- , C. H. Christensen
- , P. Christiansen
- , T. Chujo
- , C. Cicalo
- , L. Cifarelli
- , F. Cindolo
- , M. R. Ciupek
- , G. Clai
- , J. Cleymans
- , F. Colamaria
- , J. S. Colburn
- , D. Colella
- , A. Collu
- , M. Colocci
- , M. Concas
- , G. Conesa Balbastre
- , Z. Conesa del Valle
- , G. Contin
- , J. G. Contreras
- , M. L. Coquet
- , T. M. Cormier
- , P. Cortese
- , M. R. Cosentino
- , F. Costa
- , S. Costanza
- , P. Crochet
- , R. Cruz-Torres
- , E. Cuautle
- , P. Cui
- , L. Cunqueiro
- , A. Dainese
- , F. P. A. Damas

- , M. C. Danisch
- , A. Danu
- , I. Das
- , P. Das
- , P. Das
- , S. Das
- , S. Dash
- , S. De
- , A. De Caro
- , G. de Cataldo
- , L. De Cilladi
- , J. de Cuveland
- , A. De Falco
- , D. De Gruttola
- , N. De Marco
- , C. De Martin
- , S. De Pasquale
- , S. Deb
- , H. F. Degenhardt
- , K. R. Deja
- , L. Dello Stritto
- , S. Delsanto
- , W. Deng
- , P. Dhankher
- , D. Di Bari
- , A. Di Mauro
- , R. A. Diaz
- , T. Dietel
- , Y. Ding
- , R. Divia
- , D. U. Dixit
- , O. Djupsland
- , U. Dmitrieva
- , J. Do
- , A. Dobrin
- , B. Donigus
- , O. Dordic

- , A. K. Dubey
- , A. Dubla
- , S. Dudi
- , M. Dukhishyam
- , P. Dupieux
- , N. Dzalaiova
- , T. M. Eder
- , R. J. Ehlers
- , V. N. Eikeland
- , F. Eisenhut
- , D. Elia
- , B. Erasmus
- , F. Ercolessi
- , F. Erhardt
- , A. Erokhin
- , M. R. Ersdal
- , B. Espagnon
- , G. Eulisse
- , D. Evans
- , S. Evdokimov
- , L. Fabbietti
- , M. Faggin
- , J. Faivre
- , F. Fan
- , A. Fantoni
- , M. Fasel
- , P. Fecchio
- , A. Feliciello
- , G. Feofilov
- , A. Fernandez Tellez
- , A. Ferrero
- , A. Ferretti
- , V. J. G. Feuillard
- , J. Figiel
- , S. Filchagin
- , D. Finogeev
- , F. M. Fionda

- , G. Fiorenza
- , F. Flor
- , A. N. Flores
- , S. Foertsch
- , P. Foka
- , S. Fokin
- , E. Fragiacomo
- , E. Frajna
- , U. Fuchs
- , N. Funicello
- , C. Furget
- , A. Furs
- , J. J. Gaardhoje
- , M. Gagliardi
- , A. M. Gago
- , A. Gal
- , C. D. Galvan
- , P. Ganoti
- , C. Garabatos
- , J. R. A. Garcia
- , E. Garcia-Solis
- , K. Garg
- , C. Gargiulo
- , A. Garibli
- , K. Garner
- , P. Gasik
- , E. F. Gauger
- , A. Gautam
- , M. B. Gay Ducati
- , M. Germain
- , P. Ghosh
- , S. K. Ghosh
- , M. Giacalone
- , P. Gianotti
- , P. Giubellino
- , P. Giubilato
- , A. M. C. Glaenzer

- , P. Glassel
- , D. J. Q. Goh
- , V. Gonzalez
- , L. H. Gonzalez-Trueba
- , S. Gorbunov
- , M. Gorgon
- , L. Gorlich
- , S. Gotovac
- , V. Grabski
- , L. K. Graczykowski
- , L. Greiner
- , A. Grelli
- , C. Grigoras
- , V. Grigoriev
- , A. Grigoryan
- , S. Grigoryan
- , O. S. Groettvik
- , F. Grossa
- , J. F. Grosse-Oetringhaus
- , R. Grossos
- , G. G. Guardiano
- , R. Guernane
- , M. Guilbaud
- , K. Gulbrandsen
- , T. Gunji
- , A. Gupta
- , R. Gupta
- , S. P. Guzman
- , L. Gyulai
- , M. K. Habib
- , C. Hadjidakis
- , G. Halimoglu
- , H. Hamagaki
- , G. Hamar
- , M. Hamid
- , R. Hannigan
- , M. R. Haque

- , A. Harlenderova
- , J. W. Harris
- , A. Harton
- , J. A. Hasenbichler
- , H. Hassan
- , D. Hatzifotiadou
- , P. Hauer
- , L. B. Havener
- , S. Hayashi
- , S. T. Heckel
- , E. Hellbar
- , H. Helstrup
- , T. Herman
- , E. G. Hernandez
- , G. Herrera Corral
- , F. Herrmann
- , K. F. Hetland
- , H. Hillemanns
- , C. Hills
- , B. Hippolyte
- , B. Hofman
- , B. Hohlweger
- , J. Honermann
- , G. H. Hong
- , D. Horak
- , S. Hornung
- , A. Horzyk
- , R. Hosokawa
- , P. Hristov
- , C. Hughes
- , P. Huhn
- , T. J. Humanic
- , H. Hushnud
- , L. A. Husova
- , A. Hutson
- , D. Hutter
- , J. P. Iddon

- , R. Ilkaev
- , H. Ilyas
- , M. Inaba
- , G. M. Innocenti
- , M. Ippolitov
- , A. Isakov
- , M. S. Islam
- , M. Ivanov
- , V. Ivanov
- , V. Izucheev
- , M. Jablonski
- , B. Jacak
- , N. Jacazio
- , P. M. Jacobs
- , S. Jadlovska
- , J. Jadlovsky
- , S. Jaelani
- , C. Jahnke
- , M. J. Jakubowska
- , A. Jalotra
- , M. A. Janik
- , T. Janson
- , M. Jercic
- , O. Jevons
- , F. Jonas
- , P. G. Jones
- , J. M. Jowett
- , J. Jung
- , M. Jung
- , A. Junique
- , A. Jusko
- , J. Kaewjai
- , P. Kalinak
- , A. Kalweit
- , V. Kaplin
- , S. Kar
- , A. Karasu Uysal

- , D. Karatovic
- , O. Karavichev
- , T. Karavicheva
- , P. Karczmarczyk
- , E. Karpechev
- , A. Kazantsev
- , U. Kebschull
- , R. Keidel
- , D. L. D. Keijdener
- , M. Keil
- , B. Ketzer
- , Z. Khabanova
- , A. M. Khan
- , S. Khan
- , A. Khanzadeev
- , Y. Kharlov
- , A. Khatun
- , A. Khuntia
- , B. Kileng
- , B. Kim
- , C. Kim
- , D. Kim
- , D. J. Kim
- , E. J. Kim
- , J. Kim
- , J. S. Kim
- , J. Kim
- , J. Kim
- , J. Kim
- , M. Kim
- , S. Kim
- , T. Kim
- , S. Kirsch
- , I. Kisel
- , S. Kiselev
- , A. Kisiel
- , J. P. Kitowski

- , J. L. Klay
- , J. Klein
- , S. Klein
- , C. Klein-Bosing
- , M. Kleiner
- , T. Klemenz
- , A. Kluge
- , A. G. Knospe
- , C. Kobdaj
- , M. K. Kohler
- , T. Kollegger
- , A. Kondratyev
- , N. Kondratyeva
- , E. Kondratyuk
- , J. Konig
- , S. A. Konigstorfer
- , P. J. Konopka
- , G. Kornakov
- , S. D. Koryciak
- , L. Koska
- , A. Kotliarov
- , O. Kovalenko
- , V. Kovalenko
- , M. Kowalski
- , I. Kralik
- , A. Kravčakova
- , L. Kreis
- , M. Krivda
- , F. Krizek
- , K. Krizkova Gajdosova
- , M. Kroesen
- , M. Kruger
- , E. Kryshen
- , M. Krzewicki
- , V. Kučera
- , C. Kuhn
- , P. G. Kuijer

- , T. Kumaoka
- , D. Kumar
- , L. Kumar
- , N. Kumar
- , S. Kundu
- , P. Kurashvili
- , A. Kurepin
- , A. B. Kurepin
- , A. Kuryakin
- , S. Kushpil
- , J. Kvapil
- , M. J. Kweon
- , J. Y. Kwon
- , Y. Kwon
- , S. L. La Pointe
- , P. La Rocca
- , Y. S. Lai
- , A. Lakrathok
- , M. Lamanna
- , R. Langoy
- , K. Lapidus
- , P. Larionov
- , E. Laudi
- , L. Lautner
- , R. Lavicka
- , T. Lazareva
- , R. Lea
- , J. Lehrbach
- , R. C. Lemmon
- , I. Leon Monzon
- , E. D. Lesser
- , M. Lettrich
- , P. Levai
- , X. Li
- , X. L. Li
- , J. Lien
- , R. Lietava

- , B. Lim
- , S. H. Lim
- , V. Lindenstruth
- , A. Lindner
- , C. Lippmann
- , A. Liu
- , J. Liu
- , I. M. Lofnes
- , V. Loginov
- , C. Loizides
- , P. Loncar
- , J. A. Lopez
- , X. Lopez
- , E. Lopez Torres
- , J. R. Luhder
- , M. Lunardon
- , G. Luparello
- , Y. G. Ma
- , A. Maevskaya
- , M. Mager
- , T. Mahmoud
- , A. Maire
- , M. Malaev
- , N. M. Malik
- , Q. W. Malik
- , L. Malinina
- , D. Mal'Kevich
- , N. Mallick
- , P. Malzacher
- , G. Mandaglio
- , V. Manko
- , F. Manso
- , V. Manzari
- , Y. Mao
- , J. Mareš
- , G. V. Margagliotti
- , A. Margotti

- , A. Marin
- , C. Markert
- , M. Marquard
- , N. A. Martin
- , P. Martinengo
- , J. L. Martinez
- , M. I. Martinez
- , G. Martinez Garcia
- , S. Masciocchi
- , M. Masera
- , A. Masoni
- , L. Massacrier
- , A. Mastroserio
- , A. M. Mathis
- , O. Matonoha
- , P. F. T. Matuoka
- , A. Matyja
- , C. Mayer
- , A. L. Mazuecos
- , F. Mazzaschi
- , M. Mazzilli
- , M. A. Mazzoni
- , J. E. Mdhluli
- , A. F. Mechler
- , F. Meddi
- , Y. Melikyan
- , A. Menchaca-Rocha
- , E. Meninno
- , A. S. Menon
- , M. Meres
- , S. Mhlanga
- , Y. Miake
- , L. Micheletti
- , L. C. Migliorin
- , D. L. Mihaylov
- , K. Mikhaylov
- , A. N. Mishra

- , D. Miśkowiec
- , A. Modak
- , A. P. Mohanty
- , B. Mohanty
- , M. Mohisin Khan
- , Z. Moravcová
- , C. Mordasini
- , D. A. Moreira De Godoy
- , L. A. P. Moreno
- , I. Morozov
- , A. Morsch
- , T. Mrnjavac
- , V. Muccifora
- , E. Mudnic
- , D. Muhlheim
- , S. Muhuri
- , J. D. Mulligan
- , A. Mulliri
- , M. G. Munhoz
- , R. H. Munzer
- , H. Murakami
- , S. Murray
- , L. Musa
- , J. Musinsky
- , J. W. Myrcha
- , B. Naik
- , R. Nair
- , B. K. Nandi
- , R. Nania
- , E. Nappi
- , M. U. Naru
- , A. F. Nassirpour
- , A. Nath
- , C. Nattrass
- , A. Neagu
- , L. Nellen
- , S. V. Nesbo

- , G. Neskovic
- , D. Nesterov
- , B. S. Nielsen
- , S. Nikolaev
- , S. Nikulin
- , V. Nikulin
- , F. Noferini
- , S. Noh
- , P. Nomokonov
- , J. Norman
- , N. Novitzky
- , P. Nowakowski
- , A. Nyanin
- , J. Nystrand
- , M. Ogino
- , A. Ohlson
- , V. A. Okorokov
- , J. Oleniacz
- , A. C. Oliveira Da Silva
- , M. H. Oliver
- , A. Onnerstad
- , C. Oppedisano
- , A. Ortiz Velasquez
- , T. Osako
- , A. Oskarsson
- , J. Otwinowski
- , K. Oyama
- , Y. Pachmayer
- , S. Padhan
- , D. Pagano
- , G. Paić
- , A. Palasciano
- , J. Pan
- , S. Panebianco
- , P. Pareek
- , J. Park
- , J. E. Parkkila

- , S. P. Pathak
- , R. N. Patra
- , B. Paul
- , J. Pazzini
- , H. Pei
- , T. Peitzmann
- , X. Peng
- , L. G. Pereira
- , H. Pereira Da Costa
- , D. Peresunko
- , G. M. Perez
- , S. Perrin
- , Y. Pestov
- , V. Petráček
- , M. Petrovici
- , R. P. Pezzi
- , S. Piano
- , M. Pikna
- , P. Pillot
- , O. Pinazza
- , L. Pinsky
- , C. Pinto
- , S. Pisano
- , M. Płoskoń
- , M. Planinic
- , F. Pliquett
- , M. G. Poghosyan
- , B. Polichtchouk
- , S. Politano
- , N. Poljak
- , A. Pop
- , S. Porteboeuf-Houssais
- , J. Porter
- , V. Pozdniakov
- , S. K. Prasad
- , R. Preghenella
- , F. Prino

- , C. A. Pruneau
- , I. Pshenichnov
- , M. Puccio
- , S. Qiu
- , L. Quaglia
- , R. E. Quishpe
- , S. Ragoni
- , A. Rakotozafindrabe
- , L. Ramello
- , F. Rami
- , S. A. R. Ramirez
- , A. G. T. Ramos
- , T. A. Rancien
- , R. Raniwala
- , S. Raniwala
- , S. S. Rasanen
- , R. Rath
- , I. Ravasenga
- , K. F. Read
- , A. R. Redelbach
- , K. Redlich
- , A. Rehman
- , P. Reichelt
- , F. Reidt
- , H. A. Reme-ness
- , R. Renfordt
- , Z. Rescakova
- , K. Reygers
- , A. Riabov
- , V. Riabov
- , T. Richert
- , M. Richter
- , W. Riegler
- , F. Riggi
- , C. Ristea
- , S. P. Rode
- , M. Rodriguez Cahuantzi

- , K. Roed
- , R. Rogalev
- , E. Rogochaya
- , T. S. Rogoschinski
- , D. Rohr
- , D. Rohrich
- , P. F. Rojas
- , P. S. Rokita
- , F. Ronchetti
- , A. Rosano
- , E. D. Rosas
- , A. Rossi
- , A. Rotondi
- , A. Roy
- , P. Roy
- , S. Roy
- , N. Rubini
- , O. V. Rueda
- , R. Rui
- , B. Rumyantsev
- , P. G. Russek
- , A. Rustamov
- , E. Ryabinkin
- , Y. Ryabov
- , A. Rybicki
- , H. Rytkonen
- , W. Rzesz
- , O. A. M. Saarimaki
- , R. Sadek
- , S. Sadovsky
- , J. Saetre
- , K. Šafařík
- , S. K. Saha
- , S. Saha
- , B. Sahoo
- , P. Sahoo
- , R. Sahoo

- , S. Sahoo
- , D. Sahu
- , P. K. Sahu
- , J. Saini
- , S. Sakai
- , S. Sambyal
- , V. Samsonov
- , D. Sarkar
- , N. Sarkar
- , P. Sarma
- , V. M. Sarti
- , M. H. P. Sas
- , J. Schambach
- , H. S. Scheid
- , C. Schiaua
- , R. Schicker
- , A. Schmah
- , C. Schmidt
- , H. R. Schmidt
- , M. O. Schmidt
- , M. Schmidt
- , N. V. Schmidt
- , A. R. Schmier
- , R. Schotter
- , J. Schukraft
- , Y. Schutz
- , K. Schwarz
- , K. Schweda
- , G. Scioli
- , E. Scomparin
- , J. E. Seger
- , Y. Sekiguchi
- , D. Sekihata
- , I. Selyuzhenkov
- , S. Senyukov
- , J. J. Seo
- , D. Serebryakov

- , L. Šerkšnytė
- , A. Sevcenco
- , T. J. Shaba
- , A. Shabanov
- , A. Shabetai
- , R. Shahoyan
- , W. Shaikh
- , A. Shangaraev
- , A. Sharma
- , H. Sharma
- , M. Sharma
- , N. Sharma
- , S. Sharma
- , U. Sharma
- , O. Sheibani
- , K. Shigaki
- , M. Shimomura
- , S. Shirinkin
- , Q. Shou
- , Y. Sibirjak
- , S. Siddhanta
- , T. Siemiarczuk
- , T. F. Silva
- , D. Silvermyr
- , G. Simonetti
- , B. Singh
- , R. Singh
- , R. Singh
- , R. Singh
- , V. K. Singh
- , V. Singhal
- , T. Sinha
- , B. Sitar
- , M. Sitta
- , T. B. Skaali
- , G. Skorodumovs
- , M. Slupecki

- , N. Smirnov
- , R. J. M. Snellings
- , C. Soncco
- , J. Song
- , A. Songmoolnak
- , F. Soramel
- , S. Sorensen
- , I. Sputowska
- , J. Stachel
- , I. Stan
- , P. J. Steffanic
- , S. F. Stiefelmaier
- , D. Stocco
- , I. Storehaug
- , M. M. Storetvedt
- , C. P. Stylianidis
- , A. A. P. Suaide
- , T. Sugitate
- , C. Suire
- , M. Suljic
- , R. Sultanov
- , M. Šumbera
- , V. Sumberia
- , S. Sumowidagdo
- , S. Swain
- , A. Szabo
- , I. Szarka
- , U. Tabassam
- , S. F. Taghavi
- , G. Taillepied
- , J. Takahashi
- , G. J. Tambave
- , S. Tang
- , Z. Tang
- , M. Tarhini
- , M. G. Tarzila
- , A. Tauro

- , G. Tejeda Munoz
- , A. Telesca
- , L. Terlizzi
- , C. Terrevoli
- , G. Tersimonov
- , S. Thakur
- , D. Thomas
- , R. Tieulent
- , A. Tikhonov
- , A. R. Timmins
- , M. Tkacik
- , A. Toia
- , N. Topilskaya
- , M. Toppi
- , F. Torales-Acosta
- , T. Tork
- , S. R. Torres
- , A. Trifiro
- , S. Tripathy
- , T. Tripathy
- , S. Trogolo
- , G. Trombetta
- , V. Trubnikov
- , W. H. Trzaska
- , T. P. Trzcinski
- , B. A. Trzeciak
- , A. Tumkin
- , R. Turrisi
- , T. S. Tveter
- , K. Ullaland
- , A. Uras
- , M. Urioni
- , G. L. Usai
- , M. Vala
- , N. Valle
- , S. Vallero
- , N. van der Kolk

- , L. V. R. van Doremalen
- , M. van Leeuwen
- , P. Vande Vyvre
- , D. Varga
- , Z. Varga
- , M. Varga-Kofarago
- , A. Vargas
- , M. Vasileiou
- , A. Vasiliev
- , O. Vazquez Doce
- , V. Vechernin
- , E. Vercellin
- , S. Vergara Limon
- , L. Vermunt
- , R. Vertesi
- , M. Verweij
- , L. Vickovic
- , Z. Vilakazi
- , O. Villalobos Baillie
- , G. Vino
- , A. Vinogradov
- , T. Virgili
- , V. Vislavicius
- , A. Vodopyanov
- , B. Volkel
- , M. A. Volkl
- , K. Voloshin
- , S. A. Voloshin
- , G. Volpe
- , B. von Haller
- , I. Vorobyev
- , D. Voscek
- , N. Vozniuk
- , J. Vrlakova
- , B. Wagner
- , C. Wang
- , D. Wang

- , M. Weber
- , R. J. G. V. Weelden
- , A. Wegrzynek
- , S. C. Wenzel
- , J. P. Wessels
- , J. Wiechula
- , J. Wikne
- , G. Wilk
- , J. Wilkinson
- , G. A. Willems
- , B. Windelband
- , M. Winn
- , W. E. Witt
- , J. R. Wright
- , W. Wu
- , Y. Wu
- , R. Xu
- , S. Yalcin
- , Y. Yamaguchi
- , K. Yamakawa
- , S. Yang
- , S. Yano
- , Z. Yin
- , H. Yokoyama
- , I.-K. Yoo
- , J. H. Yoon
- , S. Yuan
- , A. Yuncu
- , V. Zaccolo
- , A. Zaman
- , C. Zampolli
- , H. J. C. Zanoli
- , N. Zardoshti
- , A. Zarochentsev
- , P. Zavada
- , N. Zaviyalov
- , H. Zbroszczyk

- , M. Zhalov
- , S. Zhang
- , X. Zhang
- , Y. Zhang
- , V. Zhrebchevskii
- , Y. Zhi
- , N. Zhigareva
- , D. Zhou
- , Y. Zhou
- , J. Zhu
- , Y. Zhu
- , A. Zichichi
- , G. Zinovjev
- & N. Zurlo

Contributions

All authors have contributed to the publication, being variously involved in the design and the construction of the detectors, in writing software, calibrating subsystems, operating the detectors and acquiring data, and finally analysing the processed data. The ALICE Collaboration members discussed and approved the scientific results. The manuscript was prepared by a subgroup of authors appointed by the collaboration and subject to an internal collaboration-wide review process. All authors reviewed and approved the final version of the manuscript.

Ethics declarations

Competing interests

The authors declare no competing interests.

Peer review

Peer review information

Nature thanks Benjamin Nachman 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.

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](#)

ALICE Collaboration. Direct observation of the dead-cone effect in quantum chromodynamics. *Nature* **605**, 440–446 (2022).
<https://doi.org/10.1038/s41586-022-04572-w>

- Received: 29 June 2021
- Accepted: 21 February 2022
- Published: 18 May 2022
- Issue Date: 19 May 2022
- DOI: <https://doi.org/10.1038/s41586-022-04572-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/s41586-022-04572-w>

[Download PDF](#)

- Article
- Open Access
- [Published: 18 May 2022](#)

Microcomb-driven silicon photonic systems

- [Haowen Shu](#)¹ ORCID: orcid.org/0000-0001-5311-3349,
- [Lin Chang](#) ORCID: orcid.org/0000-0001-5311-3349² ORCID: orcid.org/0000-0001-5311-3349,
- [Yuansheng Tao](#)¹ ORCID: orcid.org/0000-0001-5311-3349,
- [Bitao Shen](#)¹ ORCID: orcid.org/0000-0001-5311-3349,
- [Weiqiang Xie](#)²,
- [Ming Jin](#) ORCID: orcid.org/0000-0002-0491-4649¹,
- [Andrew Netherton](#)²,
- [Zihan Tao](#)¹,
- [Xuguang Zhang](#)¹,
- [Ruixuan Chen](#)¹,
- [Bowen Bai](#)¹,
- [Jun Qin](#)¹,
- [Shaohua Yu](#)^{1,3},
- [Xingjun Wang](#) ORCID: orcid.org/0000-0001-8206-2544^{1,3,4} & ORCID: orcid.org/0000-0001-8206-2544^{1,3,4}
- [John E. Bowers](#) ORCID: orcid.org/0000-0003-4270-8296²

Nature volume **605**, pages 457–463 (2022)

- 636 Accesses
- 29 Altmetric
- [Metrics details](#)

Subjects

- [Frequency combs](#)
- [Integrated optics](#)
- [Silicon photonics](#)

Abstract

Microcombs have sparked a surge of applications over the past decade, ranging from optical communications to metrology^{1,2,3,4}. Despite their diverse deployment, most microcomb-based systems rely on a large amount of bulky elements and equipment to fulfil their desired functions, which is complicated, expensive and power consuming. By contrast, foundry-based silicon photonics (SiPh) has had remarkable success in providing versatile functionality in a scalable and low-cost manner^{5,6,7}, but its available chip-based light sources lack the capacity for parallelization, which limits the scope of SiPh applications. Here we combine these two technologies by using a power-efficient and operationally simple aluminium-gallium-arsenide-on-insulator microcomb source to drive complementary metal–oxide–semiconductor SiPh engines. We present two important chip-scale photonic systems for optical data transmission and microwave photonics, respectively. A microcomb-based integrated photonic data link is demonstrated, based on a pulse-amplitude four-level modulation scheme with a two-terabit-per-second aggregate rate, and a highly reconfigurable microwave photonic filter with a high level of integration is constructed using a time-stretch approach. Such synergy of a microcomb and SiPh integrated components is an essential step towards the next generation of fully integrated photonic systems.

Main

Integrated photonics is profoundly impacting data communication and signal processing^{8,9,10}. A crucial development in the past decade is the demonstration of Kerr microcombs, which provide mutually coherent and equidistant optical frequency lines generated by microresonators^{1,11,12}. With a wide range of microcomb-based optoelectronic systems^{2,4,13,14,15,16,17,18}

demonstrated recently, these integrated light sources hold the promise to extend the application space of integrated photonics to a much broader scope. However, despite the tremendous progress made in microcomb integration^{19,20,21,22,23}, in almost all system-level demonstrations leveraging microcomb technologies, the passive comb generators are still the only integrated component. The rest of the system, including the comb pumping lasers, passive and active optical components, and the supporting electronics, usually rely on bulky, expensive and power-consuming equipment, thereby undermining the promised benefits of integrated photonics.

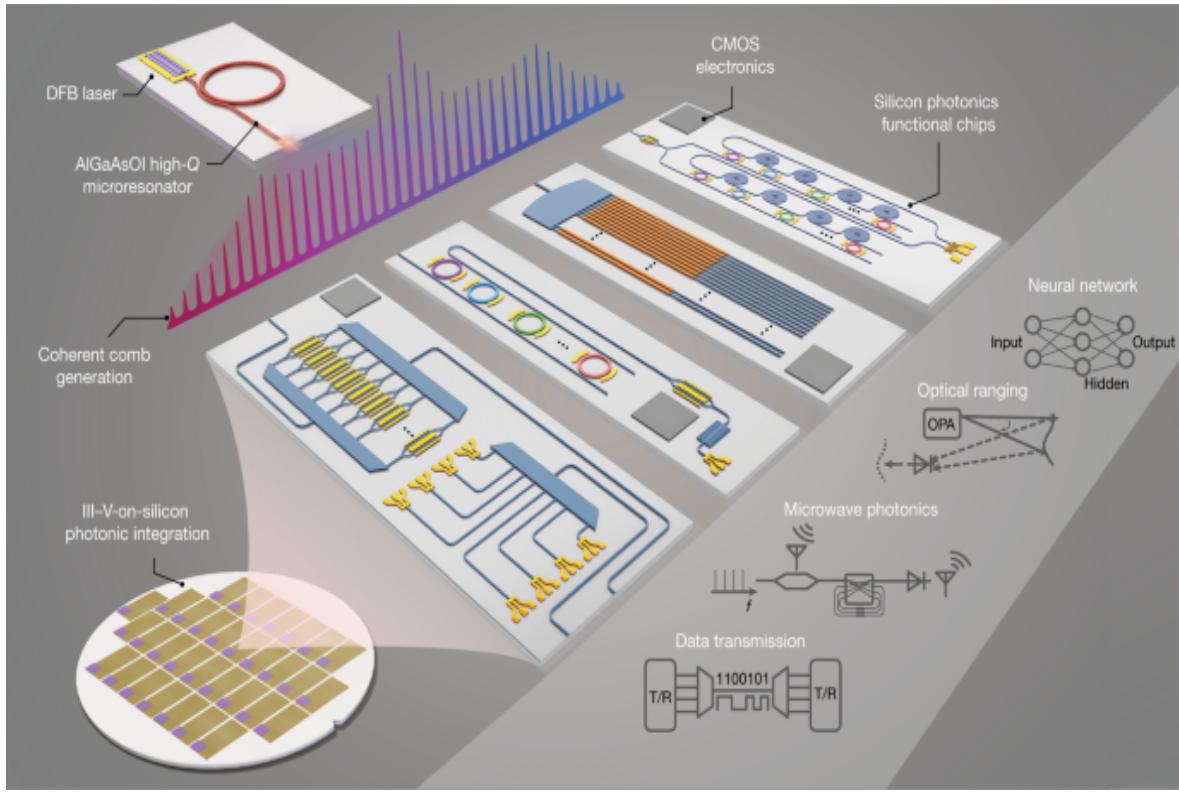
In contrast, the advances in silicon photonics (SiPh) technology have provided a scalable and low-cost solution to miniaturize optical systems^{6,24,25}, benefiting from complementary metal–oxide–semiconductor (CMOS)-compatible manufacturing. These ‘photonic engines’, have been commercialized in data interconnects^{26,27}, and widely applied in other fields^{28,29,30,31}. Yet, a key ingredient missing from foundry-based silicon-on-insulator (SOI) photonic integrated circuits (PICs) is the multiple wavelength source. For example, the current state-of-the-art photonic transceiver module contains an eight-channel distributed feedback laser (DFB) array for wavelength division multiplexing (WDM)³². Increasing the channel count in such a system requires considerable design effort, such as line-to-line spacing stabilization and increased assembly workload. Moreover, the lack of mutual coherence among channel lines restricts many applications, such as precise time–frequency metrology.

Although interfacing these two technologies is essential to address the aforementioned problems on both sides, until now, such a combination has remained elusive. Previously, although the combinations of a microcomb and other photonic components have shown potential in optical computation¹⁵, atomic clocks⁴ and synthesizer systems³, these integrated demonstrations usually rely on specialized fabrication processes unsuitable for high-volume production. Moreover, comb start-up^{33,34} and stabilization techniques^{35,36}, which require high-performance discrete optics and electronic components, markedly increase the operation complexity and system size. Recent progress in hybrid or heterogeneous laser-microcomb integration enables on-chip comb generation in a simplified manner^{21,22,23},

but these schemes add complexity in processing. These difficulties, along with the extra expenditures on multi-channel match-up and other pretreatments in system operations, have so far obstructed the implementation of a functional laser-microcomb system.

Here we make a key step in combining these two essential technologies. Using an aluminium gallium arsenide (AlGaAs)-on-insulator (AlGaAsOI) microresonator that can be directly pumped by a DFB on-chip laser, a dark-pulse microcomb is generated, which exhibits state-of-the-art efficiency, simple operation and long-time stability. Such a coherent comb is used to drive CMOS-foundry-based SiPh engines containing versatile functionality, which can be used for a wide range of applications (Fig. 1). On the basis of this approach, system-level demonstrations are presented for two major integrated photonics fields. (1) As a communications demonstration, we present a microcomb-SiPh transceiver-based data link with 100-Gbps pulse-amplitude four-level modulation (PAM4) transmission and 2-Tbps aggregate rate for data centres. (2) For microwave photonics, a compact microwave filter is demonstrated with tens-of-microseconds-level reconfiguration speed by an on-chip multitap delay-line processing scheme, whose tunable bandwidth and flexible centre frequency are capable of supporting fifth-generation (5G), radar and on-chip signal processing. This work paves the way towards the full integration of a wide range of optical systems, and will significantly accelerate the proliferation of microcombs and SiPh technologies for the next generation of integrated photonics.

Fig. 1: Microcomb-based SiPh optoelectronic systems.



Conceptual drawings for several integrated optoelectronic systems (data transmission, microwave photonic signal processing, optical beam steering and photonic computing) realized by combining a microcomb source with silicon photonic chips. With III–V-on-silicon photonic integration, the chips are expected to contain all the essential functions (for example, laser-microcomb generation, passive and active optical components, and the electronics for supporting signal processing and system control).

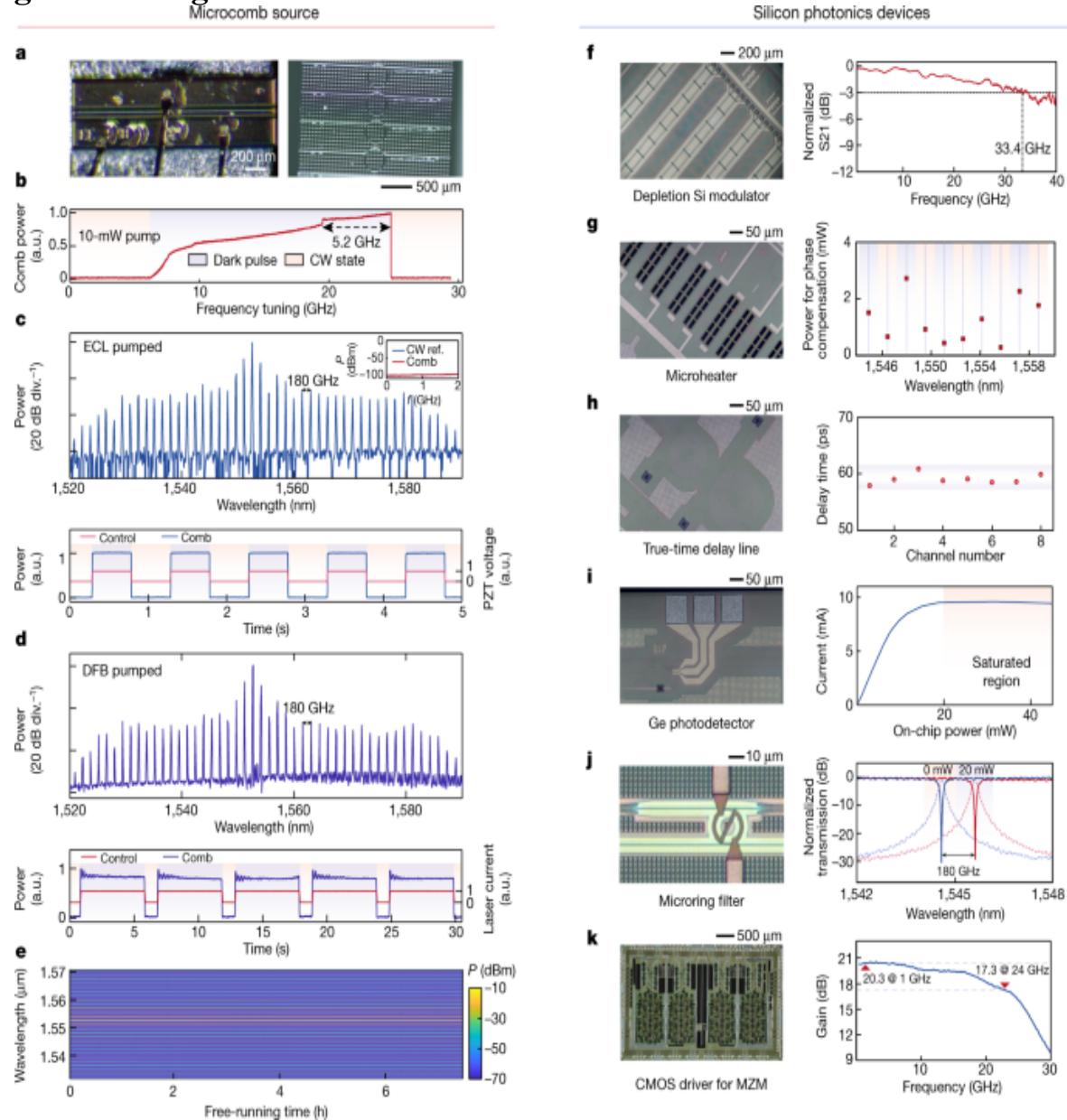
Building blocks

AlGaAsOI microcombs

The integrated comb source used in this work is based on an AlGaAsOI platform³⁷ by heterogeneous integration, as shown in Fig. 2a. Combined with the extremely high third-order nonlinear coefficient of AlGaAs ($n_2 \approx 2.6 \times 10^{-17} \text{ m}^2 \text{ W}^{-1}$), Kerr comb generation from the AlGaAsOI microresonators (Fig. 2a, right) with a moderate quality (Q) factor (one million to two million) exhibits a record-low parametric oscillation threshold

down to tens of microwatts and coherent comb-state generation under pump power at the few-milliwatts level, which can be satisfied by a commercial indium phosphide (InP) DFB laser chip (Fig. 2a, left).

Fig. 2: Comb generation and fundamental characteristics.



a, Optical image of the InP DFB laser chip and the AlGaAsOI microresonators for dark-pulse generation. **b**, Normalized comb power when tuning the pump frequency across the resonance at around 1,552 nm. With 10-mW pump power, a dark-pulse Kerr comb could be accessed in a large frequency window (tens of gigahertz). CW, continuous wave. **c, d**, Two-FSR

dark-pulse spectra (top) and the ‘turnkey’ behaviours (bottom) pumped by a commercial external laser (**c**) or a DFB laser chip (**d**) with an equal on chip power of 10 mW. A pair of flat wings besides the pump is formed in both spectra, exhibiting typical profile of the coherent dark-pulse microcombs. Inset: comb intensity noise (resolution BW of 100 kHz). The intensity noise of the dark-pulse Kerr comb is at the same power level as the electrical spectrum analyser background. P, power; f, frequency; PZT, lead zirconate titanate. **e**, Long-term stability of a free-running comb. **f–k**, Optical images and main performance of several Si-based fundamental devices, including a depletion-mode Si MZM (**f**), a TiN microheater (**g**), a Si spiral waveguide delay line (**h**), a vertical epitaxial Ge PD (**i**), a microring filter (**j**) and a CMOS driver for MZMs (**k**). More details can be found in Methods.

Besides efficiency, the operation simplicity and stability of the comb source are also critical for practical applications. In the anomalous dispersion regime, a special type of bright soliton, termed ‘soliton crystals’³⁸, exhibit these features to support system-level demonstrations without relying on electronic control^{14,16,39}. In this paper, a dark-pulse state^{40,41} is used to achieve coherent microcombs. This state works in the normal dispersion regime with the assistance of the avoided mode-crossing effect (Supplementary Note I). The dark-pulse operation experiences a much smaller power step during the transition to the coherent comb state (Supplementary Note II). More importantly, owing to the thermo-induced self-stable equilibrium mechanism of microcavities, the strong thermo-optic effects of AlGaAs ($2.3 \times 10^{-4} \text{ K}^{-1}$) here can be leveraged to significantly extend the accessibility window of the coherent comb state⁴². Such behaviour is experimentally characterized in Fig. 2b, where the comb power versus pump detuning is recorded, showing the accessible frequency range of the dark pulse to tens of gigahertz, about ten times wider than that with bright solitons³³.

Together, these traits make coherent comb generation efficient and robust in AlGaAsOI microresonators, with greatly simplified operation. Figure 2c,d shows the dark-pulse spectra pumped by an external cavity laser and a DFB laser chip, respectively, with the same on-chip power of 10 mW. Such a state can be deterministically triggered by simply turning on the laser without relying on any tuning control of electronics, thus showing ‘turnkey’

behaviour (Methods). Moreover, benefiting from the self-stabilization enabled by the strong thermo-optic effect, the comb is able to maintain stable operation without feedback loops. Figure 2d shows the spectral power versus time in a free-running AlGaAs dark pulse, with small power fluctuations over 7 h. The simplicity of both generation and stabilization facilitates seamless implementation of AlGaAsOI microcombs in current optoelectronic systems and are well suited for practical applications.

Silicon photonic engines

A monolithic SiPh circuit is used to process the generated comb lines for diverse optoelectronics systems. Such ‘silicon photonic engines’ provide functionality such as filtering, modulation, multiplexing, time delay and detection on the same chip. Figure 2fk shows the essential photonic building blocks of the optical processing engines and their key performance metrics. For signal encoding, Mach–Zehnder interferometer (MZI) travelling-wave PN depletion modulators with >33-GHz electro-optical bandwidth are used (Fig. 2f). Heaters are used to match up the modulators with the comb channels by thermal tuning (Fig. 2g). A representative result for such phase compensation in a modulator at different channel wavelengths is shown in Fig. 2g (left). To implement on-chip true-time delays, spiral waveguides with adiabatic bends are designed, as shown in Fig. 2h. The deviation of 60-ps delay lines is within 3 ps. Figure 2i shows the germanium (Ge) photodetector (PD) with about $0.5\text{--}0.8 \text{ A W}^{-1}$ at different on-chip power levels, and with a saturation power of approximately 20 mW. A microring filter array is used here to control the comb lines individually, as shown in Fig. 2j. A 180-GHz-wide (2 free spectral range (FSR)) channel-selecting range can be obtained with 20-mW heater power (Methods). In addition, the SiPh devices support system-level assembly with electronic integrated chips (Fig. 2k), allowing future integration of low-noise trans-impedance amplifiers and high-speed drivers.

System demonstrations

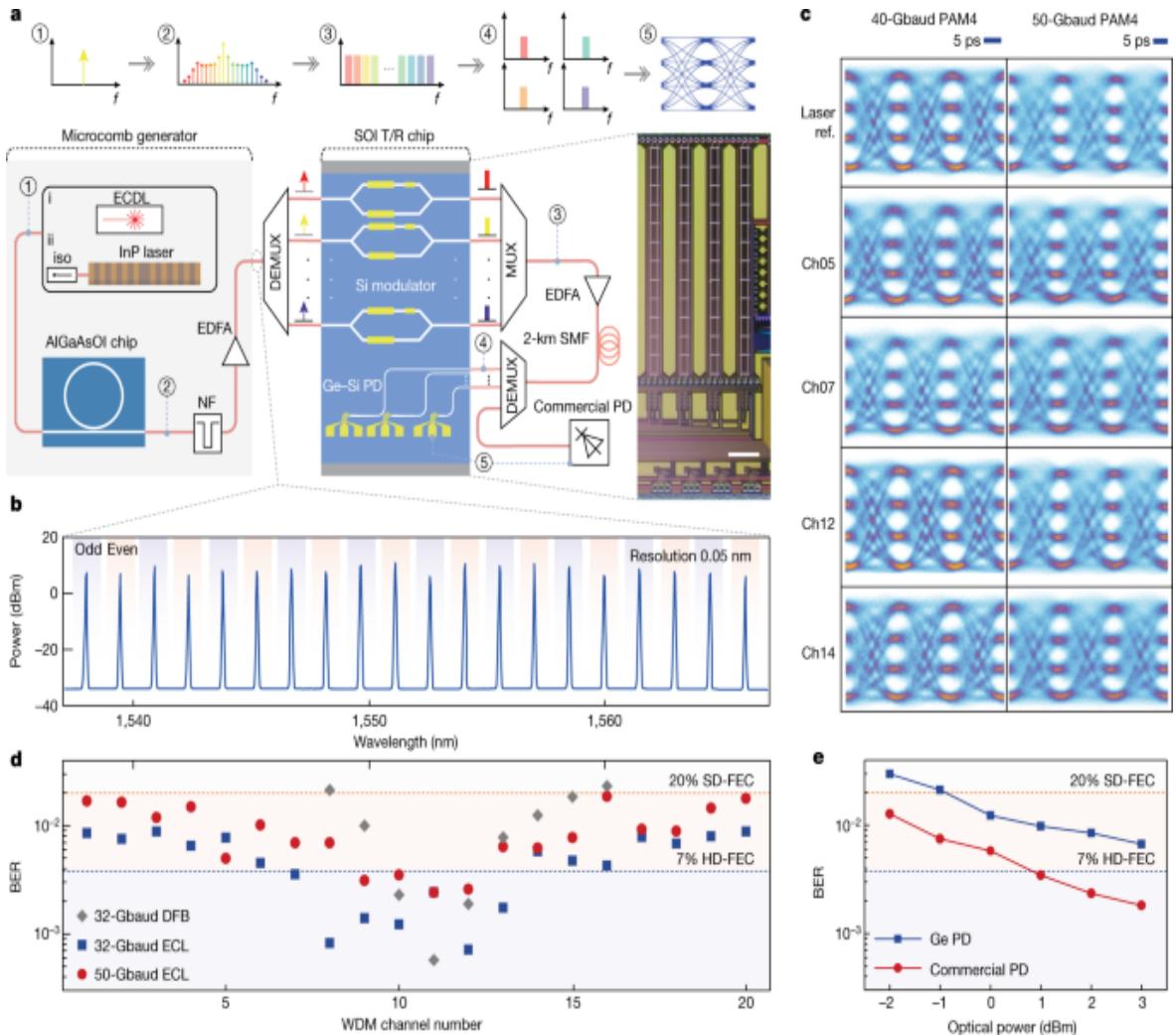
Next, two pivotal system-level demonstrations are presented: (1) a microcomb-based integrated photonic data link with a greatly increased data rate compared with traditional Si-based transceivers and (2) a rapidly

reconfigurable microcomb-based microwave photonics filter with a high level of integration.

Parallel optical data link

A schematic of the PAM4 WDM transmission system is shown in Fig. 3a. The channel spacing of the AlGaAsOI dark-pulse comb can be reconfigured from 1-FSR to multi-FSR via appropriate pre-calibration of the start-up setup (for example, laser detuning, temperature and so on)⁴¹. To achieve a higher average optical carrier-to-noise ratio while providing sufficient channel counts within the operation band, a 2-FSR spacing comb is selected as the WDM source here owing to its higher comb-line power. For pumps, a DFB laser chip and a commercially available external cavity laser (ECL) pump source are used, respectively. After the comb generation, an amplifier is needed to compensate the extra penalty brought by the demultiplexing and coupling loss. The spectrum, after amplification, is shown in Fig. 3b, in which 20 consecutive comb modes (from 1,537 nm to 1,567 nm, about 3.75-THz wide) are displayed with <5-dB power difference with proper thermal pre-setting (Methods). A simplified scheme is used to verify the chip-scale data transmission capability for carrying multi-terabits per second. The comb lines are filtered out and split into odd and even test bands by a wavelength selective switch (WSS) and then launched into the SiPh transmitting-receiving (T/R) chip, including Si modulators and Ge photodiodes. On each WDM channel, the SiPh modulators encode the carrier into PAM4 signal format at symbol rates from 32 Gbaud to 50 Gbaud. Figure 3c shows representative examples of eye diagrams after traversing 2-km-fibre links. At the receiving side, the signal is partly coupled to an on-chip Ge photodiode, whereas the remaining part is sent into a commercial PD for performance comparison. The bit-error ratio (BER) of each channel is calculated after direct detection (Methods).

Fig. 3: Transmission results.



a, Schematic of the microcomb-based data transmission set-up. The dark-pulse Kerr comb source is pumped by a continuous-wave laser, which can be generated by a commercial external cavity diode laser (ECDL, i) or a distributed feedback laser chip (ii). The generated comb is then sent into a SiPh T/R chip. iso, isolator; NF, notch filter; DEMUX, demultiplexer; MUX, multiplexer. Scale bar, 500 μ m. **b**, A 20-line comb spectrum in the C band as the multiwavelength source before injection into the SiPh T/R chip. **c**, Typical eye diagrams of the chosen channel after modulation by SiPh modulators at different symbol rates (32 Gbaud, 40 Gbaud and 50 Gbaud). **d**, BER for each comb line. The blue squares and red circles indicate the ECL-pumped comb data transmission results at symbol rates of 32 Gbaud and 50 Gbaud, respectively. All channels are considered within the given HD-FEC (3.8×10^{-3}) or SD-FEC (2×10^{-2}) threshold (blue and orange dashed lines, respectively). The grey diamond markers show the performance when

pumping the AlGaAs microresonator with a DFB chip. The wavelength-dependent BERs mainly result from the increased noise of the pre-amplifier at the edge of its operation band. The optimized receiving power for each channel is about 2–3 dBm. **e**, BER versus receiving power comparison between an on-chip Ge–Si PD and a commercial PD with the variation of the receiving power. The main limitation of the Ge–Si PD is the non-optimized frequency response (Methods).

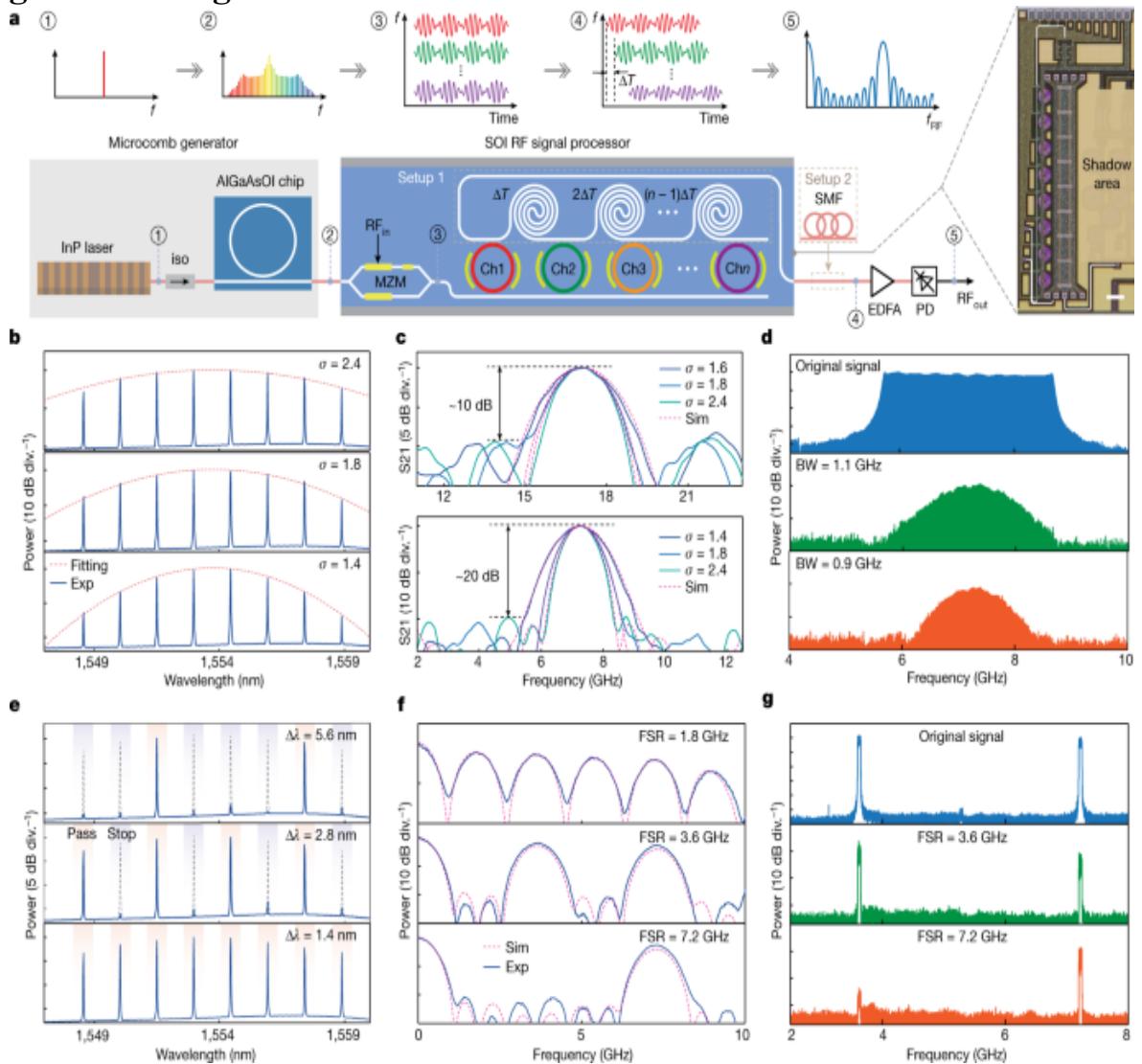
Such a dense wavelength division multiplexing scheme can greatly improve the aggregate bit rate while maintaining excellent scalability. In our proof-of-concept demonstrations, 20 comb lines in the C-band are used as the source. Figure 3d shows the BER results under three scenarios: (1) 32-Gbaud and (2) 50-Gbaud PAM4 with an ECL pump, and (3) 32-Gbaud PAM4 with a DFB pump. Considering the ECL-pumped microcomb, 7(4) channels are below the 7% hard-decision forward error correction (HD-FEC) threshold at the symbol rate of 32(50) Gbaud, with the remaining channels below the 20% soft-decision forward error correction (SD-FEC) threshold. In this case, the microcomb-based SiPh transmitter enables a baud rate of 50 Gbaud per single lane, corresponding to an aggregate bit rate of 2 Tbit s⁻¹ (1.65 Tbit s⁻¹ net rate after FEC overhead subtraction). For a higher-level integrated system, the commercial ECL pump is replaced by a DFB laser chip. With the integrated pump source, the transmitter achieves a total data transmission rate of 448 Gbit s⁻¹, with 7 channels under the FEC threshold. Another advantage of SiPh is the possibility of integrating the transmitter and receiver. BER results after optical to electrical (O/E) conversion by both commercial III–V photodiodes and on-chip Ge photodiodes are shown in Fig. 3e. At the 20% SD-FEC threshold, the penalty between two devices is approximately 2.3 dB at 32 Gbaud (Methods).

Reconfigurable microwave photonic filter

The reconfigurable microwave photonic filter (MPF) is constructed using a tapped delay line (TDL)⁴³. It is worth mentioning that TDL-based MPFs can follow two approaches depending on whether the tap delays are produced by non-dispersive (true-time) delay lines⁴⁴ or dispersive delay lines^{13,45,46,47}. In this work, both approaches are implemented. A schematic of the

experimental setup is shown in Fig. 4a. The 180-GHz-spacing microcomb served as taps for the MPF. The comb lines are then manipulated by a SiPh signal processor containing a high-speed Mach–Zehnder modulator (MZM), an eight-channel add-drop microring array (MRA) and spiral delay lines. The input radio frequency (RF) signal is loaded by the MZM. The MRA here acts as an on-chip optical spectral shaper (OSS) for the comb lines, performing spectrum slicing, line-by-line pulse shaping (weighting on taps) and spectrum recombination in sequence. A cluster of spiral waveguides offers a fixed time delay (ΔT) between adjacent taps. Finally, the processed comb lines are beaten in an off-chip fast PD to synthesize the RF filtering profiles.

Fig. 4: Reconfigurable MPF results.



a, Schematic of the setup to perform microcomb-based reconfigurable MPF. The time delays between comb lines are produced by on-chip spiral delay lines (setup 1) and dispersive propagation from a spool of SMF (setup 2). Scale bar, 200 μm . **b**, Optical spectra of Gaussian-apodization comb lines for BW programming (σ , Gaussian factor; blue, experiment (Exp); red, Gaussian fitting). **c**, RF filtering responses of the MPF with various passband BWs, based on the setup 1 (top) and setup 2 (bottom). The red dashed curves show the theoretical fitting results (Sim.) (Supplementary Note III). **d**, Proof-of-concept demonstration of RF filtering of a wideband RF signal. From top to bottom: RF spectra of original signal, signal after 1.1-GHz BW filter and signal after 0.9 GHz BW filter. **e, f**, Optical spectra (**e**) and corresponding RF responses (**f**) of the MPF with various FSRs, produced by modifying the comb line spacing and based on setup 2. $\Delta\lambda$, wavelength distance between adjacent comb lines. **g**, Proof-of-concept demonstration on RF filtering of a complex dual-channel RF signal. From top to bottom: RF spectra of original signal, signal after 3.6-GHz FSR filter and signal after 7.2-GHz FSR filter.

The system shows flexible reconstruction features in terms of passband bandwidth (BW) and RF FSR. Figure 4b depicts the optical spectra using Gaussian apodization on comb lines for passband BW reconfigurability⁴⁶. The corresponding RF filtering responses are given in Fig. 4c, with non-dispersive delay (top) and dispersive delay (bottom) configurations, respectively. The 3-dB BW of the MPF in the non-dispersive delay scheme can be continuously adjusted within a range of about 1.97–2.42 GHz by tuning the Gaussian parameter σ from 2.4 to 1.6. The main-to-sidelobe suppression ratio is about 10 dB. Better performance (>20-dB main-to-sidelobe suppression ratio) is achieved using the dispersive delay scheme, with a subgigahertz-level filtering BW tunability. The results in Fig. 4e,f show the reconfigurability of RF FSR by modifying the comb line spacing: comb line spacings of 5.6 nm, 2.8 nm and 1.4 nm result in RF filtering response FSRs of 1.8 GHz, 3.6 GHz and 7.2 GHz, respectively. In contrast with other state-of-the-art microcomb-based MPFs using either bulk OSS^{46,47} or changing soliton states¹³, this work significantly advances the degree of integration and the reconfiguration speed (about 53 μs ; Methods), which are crucial for modern wireless communications and avionic applications.

As a paradigm demonstration towards real-world applications, RF filtering on a practical microwave signal is illustrated in Fig. 4d, g. First, a broadband RF signal covering from 5.5 GHz to 9 GHz is shaped by changing the MPF BW from 0.9 GHz to 1.1 GHz, as shown in Fig. 4d, exhibiting reconfigurable passband widths. Moreover, to validate the FSR reconfigurability, a RF test signal is generated with a 50 Mb s^{-1} quadrature phase shift keying (QPSK) modulation at 3.6 GHz and 7.2 GHz, respectively (Fig. 4g). It can be observed that by setting the proper tap spacing with the on-chip OSS, the signal at 3.6 GHz could be optionally rejected.

Discussion

The performance of these systems can be further improved by optimizing the integrated devices or employing superior signal-processing techniques. Additional multiplexing techniques (such as space-division multiplexing and polarization-division multiplexing) and higher modulation formats (such as, PAM6 and PAM8) could be used to boost the transmission capacity. The data rate can be further scaled up to $>10 \text{ Tbps}$ by broadening the operation wavelength to the L band and the S band. The performance of the DFB-pumped integrated comb source is mainly limited by the relatively high noise floor of the free-running DFB laser (Methods), which lowers the optical signal-to-noise ratio (OSNR). For the RF filter, a narrower filtering BW (down to subgigahertz) and a higher tuning resolution can be obtained by increasing the number of tap channels used in the finite impulse response configurations⁴³, that is, expansion of the MRA.

We expect more integrated functionality be incorporated in the future, culminating in fully integrated microcomb-based optoelectronic systems. For instance, self-injection locked dark-pulse microcomb sources²¹ could be monolithically realized by using heterogeneously integrated III–V lasers and microresonators²³. The discrete erbium-doped fibre amplifiers (EDFAs) could be replaced by on-chip SOAs, which can potentially be integrated with other photonic components on the same chip^{48,49}. More recently, AlGaAs-on-SOI photonic circuits have been developed to integrate the two material platforms we used in this work on the same wafer⁵⁰. The photonic elements can also be combined with application-specific electronic circuits, which will further improve the compactness and power efficiency. Considering the

versatility offered by the technologies, microcomb-driven SiPh systems will provide a mass-produced and low-cost solution to a broad range of optoelectronics applications, therefore facilitating the next generation of integrated photonics.

Methods

Design and fabrication of the devices

The ring waveguides of the AlGaAsOI resonators were designed to work within the normal dispersion regime in the C band, with dimensions of 400 nm × 1,000 nm. The width of the bus waveguide at the facet was designed to be 200 nm for efficient chip-to-fibre coupling. The fabrication of the AlGaAs microresonators was based on heterogeneous wafer bonding technology. The process is currently realized at the 100-mm-wafer scale without any strict fabrication processes such as chemical-mechanical polishing or high-temperature annealing that are not compatible with the CMOS process. It can therefore be directly adopted by current III–V/Si photonic foundries⁵¹. A Q factor >2 million can be obtained in the AlGaAsOI resonator, corresponding to a waveguide loss of $<0.3 \text{ dB cm}^{-1}$. The fraction of aluminium is 0.2, which corresponds to a two-photon absorption wavelength of around 1,480 nm. The epitaxial wafer growth was accomplished using molecular-beam epitaxy. A 248-nm deep-ultraviolet stepper was used for the lithography. A photoresist reflow process and an optimized dry etch process were applied in waveguide patterning to minimize waveguide scattering loss. More fabrication details can be found in refs. ^{52,53}. The SiPh PIC, including its Si modulators and Si–Ge PDs, was fabricated on a 200-mm SOI wafer with a Si-layer thickness of 220 nm and a buried oxide layer thickness of 2 μm using CMOS-compatible processes at CompoundTek Pte in a one-to-one 200-mm-wafer run with its standard 90-nm lithography SOI process. The waveguide loss in this SiPh platform is approximately 1.2 dB cm^{-1} in the C band. In our experiment, lensed fibres with different mode field diameters were selected for the AlGaAsOI and SOI chips; the coupling loss is about 3–5 dB per facet for AlGaAsOI waveguides and about 2–3 dB per facet for Si waveguides.

Characterizations of the building-block units

The FSR of the 144- μm -radius rings utilized in this study is about 90 GHz. The microcomb shows advances both in start-up and stabilization. During the dark-pulse generation, a much smaller abrupt power change occurs when the comb transits from continuous-wave states to dark-pulse states, indicating the elimination of the well known triggering problem in bright soliton generation. Compared with general bright solitons, the dark pulse is inherently tolerant to thermal effects that usually make bright soliton states difficult to access⁵⁴. For long-term stability measurement, the comb spectra and comb line power of a free-running dark-pulse comb are recorded by a high-resolution optical spectrum analyser (OSA) every 5 min.

More details are presented here for the SiPh devices shown in Fig. 2. The opto-electrical BW of the depletion-mode Si-based MZMs was measured by a vector network analyser (Keysight N524), with the typical results of >30 GHz. The on-chip phase compensation units are MZI-based titanium nitride (TiN) microheaters. The resistance is approximately $200\ \Omega$. The TiN metal layer is about 1 μm above the Si layer, ensuring a heating efficiency of about $20\ \text{mW}\ \pi^{-1}$. Meanwhile, a deep trench process is utilized to isolate each microheater to diminish thermal cross-talk. For the on-chip true-time delay line, we adopted a 2- μm -wide multimode Si waveguide for low-loss transmission. Euler curves were used in the spiral waveguide for adiabatic bending. For a 60-ps Si delay line, the total loss is $<0.5\ \text{dB}$, with a delay-time variation of $<3\%$ among 8 tested devices. For the vertical epitaxial Ge PD, the responsivity declines with the increasing on-chip power. A saturated point of about 20 mW could be reached when the power is further increased. Microring filters employed for WDM could be tuned by microheaters, with which a 180-GHz channel spacing can be obtained under 20-mW power dissipation. The CMOS drivers for signal amplification before injection into the Si MZM (not used in the high-bit-rate ($>50\ \text{Gbps}$) signal transmission experiment) show a 3-dB gain BW of about 24 GHz.

The performance of other building-block devices is presented in Extended Data Fig. 1. The linewidth of the DFB laser used as the pump is measured by a delayed self-heterodyne method⁵⁵. The measurement and Lorentzian fitting result are shown in Extended Data Fig. 1a, exhibiting a laser

linewidth of about 150 kHz. For the SiPh devices, the 3-dB BW of the Si–Ge photodiodes is shown in Extended Data Fig. 1b, indicating an approximately 30-GHz S21 parameter. Such a non-optimized BW accounts for the penalty in Fig. 3c. Structure design for a lower resistor-capacitor time constant could further increase the operation BW. For on-chip monitoring, the asymmetric MMI-based 10:90 power splitter⁵⁶ is employed in the system, as shown in Extended Data Fig. 1c. The symmetry of the multimode region is broken by removing the corner of the MMI (marked with a red dashed rectangle), which causes a dramatic redistribution of the optical field, thus leading to an uneven power splitting by changing the width of the cut-off corner. We randomly chose four identical 1:9 MMIs and tested the power splitting ratios. The results were found to be close to the design target (dashed line), exhibiting good consistency, as shown in the bottom panel of Extended Data Fig. 1c. Moreover, the grating couplers used in this work (Extended Data Fig. 1d) show a roughly 2-dB coupling efficiency difference across the operation band (1,535–1,565 nm).

Turnkey dark-pulse microcomb generation

The turnkey microcomb generation test setup is shown in Extended Data Fig. 2a, with either an ECL or a DFB laser as the pump. Slow laser-frequency detuning is enough for microcomb generation, which can be realized by adjusting the cavity length via tuning the lead zirconate titanate voltage of the commercial ECL or changing the laser current of DFB, respectively. After the comb generation, the spectra are recorded; meanwhile, the total power of the generated comb lines is measured in real time. A pre-calibration process is required to ensure the laser frequency will locate at the comb accessing range ultimately. For the ECL-pumped dark-pulse comb (Extended Data Fig. 2b), a 1-Hz square wave is used as the trigger signal, which tunes the pump wavelength about 0.3 nm away from or into the resonance. For the DFB-pumped experiments (Extended Data Fig. 2c), when a laser is turned on, there is always an automatic frequency ramping-up process owing to the injected carrier and the warming of the cavity, which can directly initiate the microcomb generation as long as the lasing frequency of the final stable state lies within the range of the access window of the coherent state. In our experiment, the laser current is switched between two values with a period of 6 s (1 s for the ‘off’ state and

5 s for the ‘on’ state). Both results show immediate on–off behaviours of microcomb generation along with the low-speed control signal. It is noted that there is some power ripple of the DFB-pumped comb in the first few seconds, which is due to the temperature vibration caused by thermoelectric cooler, after which the comb state is stabilized. The comb is reproducible in several consecutive switching tests, with great robustness.

Details of data transmission experiments

In our experiment, the microcomb is first pumped by a commercial tunable laser (Toptica CTL 1550), then by a DFB laser chip for a higher degree of integration, where an optical isolator is deployed between the DFB laser and the AlGaAsOI microresonator to eliminate the reflection. When tuning the pumping wavelength from the blue side to a certain detuned value at around 1,552.5 nm, both configurations generate dark pulses with 2-FSR comb spacing. The detailed experimental setup for data transmission is shown in Extended Data Fig. [3a](#). For the comb spectrum with large power fluctuations, an additional amplification process is required owing to the insufficient gain of those low-power channels, which introduces extra system complexity and power consumption on the transmitting side. In this work, owing to the strong thermal effect, the avoided mode-crossing (AMX) strength of the AlGaAs microresonator can be thermally pre-set to obtain a coherent microcomb with a less disparate power distribution across the operation band. Thus, only a notch filter is required to attenuate the central three comb lines for the subsequent equalized comb amplification. The comb is amplified by an EDFA and then split into odd and even test bands^{[39,57,58](#)} by a wavelength-selective switch (Finisar Waveshaper 4000s). A Si modulator and a lithium niobate (LN) modulator (EOspace, 35-GHz BW) are deployed at the odd and even bands, respectively. Ten comb lines in each test band are simultaneously modulated. The modulators are driven at a 32-Gbaud or 50-Gbaud symbol rate. The differential PAM-4 signal is generated by a commercial pulse pattern generator (Anritsu PAM4 PPG MU196020A). The insertion loss of the SiPh (LN) modulator is 13(8) dB. The SiPh modulator undergoes a relatively high loss (including the edge coupling loss of about 2 dB per facet), which results in a power difference between the two test bands. The modulated test bands are then combined by a 50:50 power coupler and launched into another WSS for comb power

equalization. At the receiving side, each WDM channel encoded by the Si modulator is sequentially filtered out and measured. Eye diagrams are produced by a sampling oscilloscope (Anritsu MP 2110A) with a 13-tap transmitter and dispersion eye closure quaternary (TDECQ) equalizer (accumulation time, 8 s). The BERs are measured online by an error detector (Anritsu PAM4 ED MU196040B) with 1-dB low-frequency equalization and a decision-feedback equalization. Extended Data Fig. [3b](#) shows the 100-Gbps PAM4 eye diagrams for each of the 20 channels.

It is worth noting that the performance is underestimated. In our proof-of-concept test configuration, ten channels in each test band are modulated at the same time. Considering two-photon absorption in Si waveguides, the maximum input power for the Si modulator is about 13 dBm, which results in only 3-dBm optical power per single lane. Moreover, considering the extra penalty introduced by the WSS for power equalization, unnecessary in real-word transmission scenarios, the OSNR for each channel can be at least 10 dB higher. Thus, a better transmission result is attainable.

Noise analysis of different pump schemes

The noise floor of the DFB and the ECL are roughly characterized in an OSA, as shown in Extended Data Fig. [4a](#). The laser spectra indicate that the noise of the DFB is evidently higher than that of the ECL. The combs in our experiments are pumped by the free-running DFB laser and the ECL separately, as shown in Extended Data Fig. [4b,c](#). With the almost same pumping power of about 10 mW, the DFB chip holds a 10-dB-higher noise floor compared with the ECL, corresponding to an equivalent OSNR reduction in each comb line. Moreover, the amplification after the comb generation would also result in OSNR degradation, which could be a potential problem when replacing the current EDFA with integrated SOAs (about 4–5-dB-noise-floor increment in a commercial EDFA and about 7 dB in commercial on-chip SOAs). The OSNR of the DFB-pumped microcomb can be further improved by employing an on-chip optical filter for comb distillation^{[59,60](#)} or introducing optical injection locking between the microcomb and slave lasers for low-noise amplification^{[61](#)}. Also, increasing the pump power will lead to a higher average OSNR and more stable long-

term behaviour, which is an advantage over the injection-locking-based dark-pulse generation^{21,62}.

Setup of the dispersive delay-line MPF scheme

As the non-uniformity of delays owing to the inevitable fabrication errors will degrade the filtering performance, the second TDL-MPF approach is also implemented to further determine the optimal filtering performance: a spool of single-mode fibre (SMF) is used instead of the on-chip spiral delay lines to produce dispersive delay. Extended Data Fig. 5 shows the experimental setup of the reconfigurable MPF carried out in a dispersive delay-line configuration. Compared with Fig. 4a, most of the MPF system remains unchanged and has one main difference, which is that the on-chip true-time spiral delay lines are removed from the SiPh signal processor. The processed comb lines will propagate through a spool of 5-km SMF (as a dispersive element) to obtain a solid delay unit between adjacent taps, which can be expressed as $T = \delta\lambda DL$ (ignoring the high-order dispersion of SMF), where $\delta\lambda$ represents the comb line spacing, D is the dispersion coefficient of SMF and the L is the length of SMF. In this scheme, the basic delay T among comb lines is generated by a single dispersive element, which can be kept as uniform value and not influenced by fabrication errors. Besides, this system is more flexible; for instance, the centre frequency of the filtering passband can be adjusted by simply change the length or dispersion coefficient of SMF.

Details of RF filter experiments

The DFB-driven dark-pulse Kerr comb exhibits 2-FSR (180-GHz) comb spacing. The initial comb source is amplified by an EDFA, and 8 comb lines in the range of 1,547–1,560 nm are selected using an optical bandpass filter before injection into a SiPh signal processor chip. The input and output coupling are achieved via grating couplers of about 40% coupling efficiency. Frequency-swept RF signals with 9-dBm power from a vector network analyzer are applied to the Si MZM in double-sideband format. The tap weighting coefficients are set by adjusting the relative detuning among the comb lines and their corresponding resonance wavelengths in the Si MRA with TiN microheaters placed on the waveguides. The output light of the Si

chip is split by a 10:90 optical power coupler: 10% of the light is sent into an optical spectrum analyser (Yokogawa AQ6370C) for spectral monitoring, whereas the other 90% of the light propagates through the follow-up optical link. In the dispersive delay scheme, a spool of 5-km SMF is used to acquire the dispersive delay between adjacent comb lines (taps). Finally, the processed comb lines are beat in a 50-GHz PD (Finisar 2150R) to convert the optical signal into electrical domain. A low-noise EDFA is placed before the PD to compensate for the link insertion loss and coupling loss.

For the practical demonstrations of RF signal filtering, a 50 Gsamples s⁻¹ arbitrary waveform generator (AWG, Tektronix AWG70001) is used to produce the desired RF input signals. To validate the BW reconfigurability of this filter, an ultrawideband RF signal is generated, spanning from 5.5 GHz to 9 GHz. To validate the FSR reconfigurability of this filter, a complex RF signal is produced that contains a 50-Mb·s⁻¹ QPSK spectrum modulated at 3.6 GHz and a 50-Mb·s⁻¹ QPSK spectrum modulated at 7.2 GHz. The RF outputs from the AWG are amplified by a linear electrical driver (SHF 807C) before routing to the Si MZM. The filtered RF signals are detected by a signal analyser (Keysight N9010B) for spectrum measurement. A similar FSR multiplication of the MPF has been reported previously and explained by temporal Talbot effects⁶³. However, the crucial Talbot processor used in these MPF systems is based on more complex discrete devices, which will increase the power dissipation and make the system less stable.

Unlike the conventional waveshaper based on bulky liquid-crystal spatial light modulators⁶⁴, one of the remarkable advantages of the chip-scale add-drop microring resonator (MRR) array used in our work is the rapid reconfiguration of RF filtering responses. The reconfiguration operation on filtering spectra is realized by adjusting the shaping profiles of comb lines, through the TiN microheater placed on the waveguides. To explore the maximum reconfiguration speed, a standard electrical square-wave waveform is generated by a function waveform generator (RIGOL, DG2102) to drive a single MRR channel. The output of the MRR is received by a photodetector (Thorlabs DET08CFC/M), and then recorded by a digital oscilloscope (RIGOL, DS7014 10 GSa s⁻¹). Extended Data Fig. 6 shows the measured switching temporal response. As seen in Extended Data Fig. 6b,c, the 90/10 rise and fall times are 15 μ s and 53 μ s, respectively. Therefore, the

fastest response speed for the reconfiguration operation is approximately 19 kHz.

Data availability

The data that supports the plots within this paper and other findings of this study are available on Zenodo (<https://doi.org/10.5281/zenodo.6092678>). All other data used in this study are available from the corresponding authors upon reasonable request.

Code availability

The codes that support the findings of this study are available from the corresponding authors upon reasonable request.

References

1. Kippenberg, T. J., Gaeta, A. L., Lipson, M. & Gorodetsky, M. L. Dissipative Kerr solitons in optical microresonators. *Science* **361**, eaan8083 (2018).
2. Spencer, D. T. et al. An optical-frequency synthesizer using integrated photonics. *Nature* **557**, 81–85 (2018).
3. Marin-Palomo, P. et al. Microresonator-based solitons for massively parallel coherent optical communications. *Nature* **546**, 274–279 (2017).
4. Newman, Z. L. et al. Architecture for the photonic integration of an optical atomic clock. *Optica* **6**, 680–685 (2019).
5. Rickman, A. The commercialization of silicon photonics. *Nat. Photon.* **8**, 579–582 (2014).
6. Bogaerts, W. & Chrostowski, L. Silicon photonics circuit design: methods, tools and challenges. *Laser Photon. Rev.* **12**, 1700237 (2018).

7. Thomson, D. et al. Roadmap on silicon photonics. *J. Opt.* **18**, 073003 (2016).
8. Chen, X. et al. The emergence of silicon photonics as a flexible technology platform. *Proc. IEEE* **106**, 2101–2116 (2018).
9. Marpaung, D., Yao, J. & Capmany, J. Integrated microwave photonics. *Nat. Photon.* **13**, 80–90 (2019).
10. Miller, D. A. B. Attojoule optoelectronics for low-energy information processing and communications. *J. Lightwave Technol.* **35**, 346–396 (2017).
11. Gaeta, A. L., Lipson, M. & Kippenberg, T. J. Photonic-chip-based frequency combs. *Nat. Photon.* **13**, 158–169 (2019).
12. Chang, L., Liu, S. & Bowers, J. E. Integrated optical frequency comb technologies. *Nat. Photon.* **16**, 95–108 (2022).
13. Hu, J. et al. Reconfigurable radiofrequency filters based on versatile soliton microcombs. *Nat. Commun.* **11**, 4377 (2020).
14. Xu, X. et al. Photonic microwave true time delays for phased array antennas using a 49 GHz FSR integrated optical microcomb source. *Photon. Res.* **6**, B30–B36 (2018).
15. Feldmann, J. et al. Parallel convolutional processing using an integrated photonic tensor core. *Nature* **589**, 52–58 (2021).
16. Xu, X. et al. 11 TOPS photonic convolutional accelerator for optical neural networks. *Nature* **589**, 44–51 (2021).
17. Yang, Q.-F. et al. Vernier spectrometer using counter propagating soliton microcombs. *Science* **363**, 965–968 (2019).
18. Riemensberger, J. et al. Massively parallel coherent laser ranging using a soliton microcomb. *Nature* **581**, 164–170 (2020).

19. Stern, B., Ji, X., Okawachi, Y., Gaeta, A. L. & Lipson, M. Battery-operated integrated frequency comb generator. *Nature* **562**, 401–405 (2018).
20. Voloshin, A. S. et al. Dynamics of soliton self-injection locking in optical microresonators. *Nat. Commun.* **12**, 235 (2021).
21. Jin, W. et al. Hertz-linewidth semiconductor lasers using cmos-ready ultra-high- Q microresonators. *Nat. Photon.* **15**, 346–353 (2021).
22. Shen, B. et al. Integrated turnkey soliton microcombs. *Nature* **582**, 365–369 (2020).
23. Xiang, C. et al. Laser soliton microcombs heterogeneously integrated on silicon. *Science* **373**, 99–103 (2021).
24. Shen, Y. et al. Silicon photonics for extreme scale systems. *J. Lightwave Technol.* **37**, 245–259 (2019).
25. Margalit, N. et al. Perspective on the future of silicon photonics and electronics. *Appl. Phys. Lett.* **118**, 220501 (2021).
26. Radhakrishnan, N. et al. Silicon photonics-based 100Gbit/s, PAM4, DWDM data center interconnects. *J. Opt. Commun. Netw.* **10**, 25–36 (2018).
27. Xie, C. et al. Real-time demonstration of silicon-photonics-based qsfp-dd 400GBASE-DR4 transceivers for datacenter applications. In *2020 Optical Fiber Communications Conference and Exhibition (OFC)* OSA Technical Digest T3H.5 (Optica Publishing Group, 2020).
28. Shen, Y. et al. Deep learning with coherent nanophotonic circuits. *Nat. Photon.* **11**, 441–446 (2017).
29. Sun, J., Timurdogan, E., Yaacobi, A., Hosseini, E. S. & Watts, M. R. Large-scale nanophotonic phased array. *Nature* **493**, 195–199 (2013).
30. Wang, J. et al. Reconfigurable radio-frequency arbitrary waveforms synthesized in a silicon photonic chip. *Nat. Commun.* **6**, 5957 (2015).

31. Wang, J. et al. Multidimensional quantum entanglement with large-scale integrated optics. *Science* **360**, 285–291 (2018).
32. Finisar 400g optical transceiver product. *EPS Global* <https://www.epsglobal.com/products/transceivers/cfp-cfp2-cfp4/400g-cfp8-lr8-1310nm-10km-lc-transceiver> (2021).
33. Moille, G. et al. Dissipative Kerr solitons in a III–V microresonator. *Laser Photon. Rev.* **14**, 2000022 (2020).
34. Zhou, H. et al. Soliton bursts and deterministic dissipative Kerr soliton generation in auxiliary-assisted microcavities. *Light Sci. Appl.* **8**, 50 (2019).
35. Geng, Y. et al. Enhancing the long-term stability of dissipative Kerr soliton microcomb. *Opt. Lett.* **45**, 5073–5076 (2020).
36. Yi, X., Yang, Q.-F., Youl Yang, K. & Vahala, K. Active capture and stabilization of temporal solitons in microresonators. *Opt. Lett.* **41**, 2037–2040 (2016).
37. Pu, M., Ottaviano, L., Semenova, E. & Yvind, K. Efficient frequency comb generation in AlGaAs-on-insulator. *Optica* **3**, 823–826 (2016).
38. Cole, D. C., Lamb, E. S., Del’Haye, P., Diddams, S. A. & Papp, S. B. Soliton crystals in Kerr resonators. *Nat. Photon.* **11**, 671–676 (2017).
39. Corcoran, B. et al. Ultra-dense optical data transmission over standard fibre with a single chip source. *Nat. Commun.* **11**, 2568 (2020).
40. Xue, X. et al. Mode-locked dark pulse Kerr combs in normal-dispersion microresonators. *Nat. Photon.* **9**, 594–600 (2015).
41. Shu, H. et al. Sub-milliwatt-level coherent microcomb generation in AlGaAs-on-insulator microresonators. Preprint at <https://arxiv.org/abs/2112.08904> (2021).
42. Carmon, T., Yang, L. & Vahala, K. J. Dynamical thermal behavior and thermal self-stability of microcavities. *Opt. Express* **12**, 4742–4750

(2004).

43. Capmany, J., Ortega, B. & Pastor, D. A tutorial on microwave photonic filters. *J. Lightwave Technol.* **24**, 201–229 (2006).
44. Liao, S. et al. Arbitrary waveform generator and differentiator employing an integrated optical pulse shaper. *Opt. Express* **23**, 12161–12173 (2015).
45. Sancho, J. et al. Integrable microwave filter based on a photonic crystal delay line. *Nat. Commun.* **3**, 1075 (2012).
46. Xu, X. et al. Advanced adaptive photonic RF filters with 80 taps based on an integrated optical micro-comb source. *J. Lightwave Technol.* **37**, 1288–1295 (2019).
47. Xue, X. et al. Programmable single-bandpass photonic RF filter based on Kerr comb from a microring. *J. Lightwave Technol.* **32**, 3557–3565 (2014).
48. Tran, M. et al. Extending the spectrum of fully integrated photonics. Preprint at <https://arxiv.org/abs/2112.02923> (2021).
49. Davenport, M. L. et al. Heterogeneous silicon/III–V semiconductor optical amplifiers. *IEEE J. Sel. Top. Quantum Electron.* **22**, 78–88 (2016).
50. Xie, W. et al. Silicon-integrated nonlinear III–V photonics. *Photon. Res.* **10**, 535–541 (2022).
51. Liang, D. & Bowers, J. E. Recent progress in heterogeneous III–V-on-silicon photonic integration. *Light Adv. Manuf.* **2**, 5 (2021).
52. Chang, L. et al. Ultra-efficient frequency comb generation in AlGaAs-on-insulator microresonators. *Nat. Commun.* **11**, 1331 (2020).
53. Xie, W. et al. Ultrahigh- Q AlGaAs-on-insulator microresonators for integrated nonlinear photonics. *Opt. Express* **28**, 32894–32906 (2020).

54. Li, Q. et al. Stably accessing octave-spanning microresonator frequency combs in the soliton regime. *Optica* **4**, 193–203 (2017).
55. Komljenovic, T. et al. Widely tunable narrow-linewidth monolithically integrated external-cavity semiconductor lasers. *IEEE J. Sel. Top. Quantum Electron.* **21**, 214 (2015).
56. Deng, Q., Liu, L., Li, X. & Zhou, Z. Arbitrary-ratio 1×2 power splitter based on asymmetric multimode interference. *Opt. Lett.* **39**, 5590 (2014).
57. Hu, H. et al. Single-source chip-based frequency comb enabling extreme parallel data transmission. *Nat. Photon.* **12**, 469–473 (2018).
58. Fülöp, A. et al. High-order coherent communications using mode-locked dark-pulse kerr combs from microresonators. *Nat. Commun.* **9**, 1598 (2018).
59. Corcoran, B. et al. Overcoming low-power limitations on optical frequency combs using a micro-ring resonator. In *2020 Optical Fiber Communications Conference and Exhibition (OFC)* OSA Technical Digest T4G.5 (Optica Publishing Group, 2020).
60. Prayoonyong, C. et al. Frequency comb distillation for optical superchannel transmission. *J. Lightwave Technol.* **132**, 7383–7392 (2021).
61. Kuse, N. & Minoshima, K. Amplification and phase noise transfer of a Kerr microresonator soliton comb for low phase noise THz generation with a high signal-to-noise ratio. *Opt. Express* **30**, 318–325 (2022).
62. Lihachev, G. et al. Platicon microcomb generation using laser self-injection locking. *Nat. Commun.* **13**, 1771 (2022).
63. Maram, R., Onori, D., Azaña, J. & Chen, L. R. Discretely programmable microwave photonic filter based on temporal talbot effects. *Opt. Express* **27**, 14381–14391 (2019).

64. Ferdous, F. et al. Spectral line-by-line pulse shaping of on-chip microresonator frequency combs. *Nat. Photon.* **5**, 770–776 (2011).

Acknowledgements

We thank Shenzhen PhotonX Technology Co., Ltd, for laser packaging support, J. Wang and W. Zou for help in layout design, J. Shi for discussion and T. J. Morin for commentary on the manuscript. The UCSB nanofabrication facility was used.

Author information

Author notes

1. These authors contributed equally: Haowen Shu, Lin Chang, Yuansheng Tao, Bitao Shen

Affiliations

1. State Key Laboratory of Advanced Optical Communications System and Networks, School of Electronics, Peking University, Beijing, China

Haowen Shu, Yuansheng Tao, Bitao Shen, Ming Jin, Zihan Tao, Xuguang Zhang, Ruixuan Chen, Bowen Bai, Jun Qin, Shaohua Yu & Xingjun Wang

2. Department of Electrical and Computer Engineering, University of California Santa Barbara, Santa Barbara, CA, USA

Lin Chang, Weiqiang Xie, Andrew Netherton & John E. Bowers

3. Peng Cheng Laboratory, Shenzhen, China

Shaohua Yu & Xingjun Wang

4. Frontiers Science Center for Nano-optoelectronics, Peking University, Beijing, China

Xingjun Wang

Contributions

The experiments were conceived by H.S., L.C., Y.T. and B.S. The devices were designed by H.S., L.C. and Y.T. The microcomb simulation and modelling was conducted by B.S. The system-level experiments were performed by H.S. and Y.T., with assistance from L.C., B.S., M.J., Z.T., X.Z., Q.J., R.C. and B.B. The AlGaAsOI microresonators were fabricated by W.X. and L.C. The results were analysed by H.S., Y.T., B.S. and A.N. All authors participated in the writing of the manuscript. The project was supervised by L.C., S.Y., X.W. and J.E.B.

Corresponding authors

Correspondence to [Xingjun Wang](#) or [John E. Bowers](#).

Ethics declarations

Competing interests

The authors declare no competing interests.

Peer review

Peer review information

Nature thanks David Moss 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 Performance of building block devices.

a, Measured linewidth of the DFB laser. **b**, The measured 3dB bandwidth of the Si-Ge PD photodiode. **c**, Design, fabrication and measurement results of the 10:90 asymmetric MMI as on-chip monitor for silicon photonics engines. **d**, Optical image of the grating coupler (left) and its normalized transmission.

Extended Data Fig. 2 Turnkey AlGaAs dark-pulse microcomb generation.

a, Experimental setup. **b**, ECL and **c**, DFB laser chip driven comb spectra and the comb power variations along with the control signal in five consecutive switching tests.

Extended Data Fig. 3 Detailed information for data transmission.

a, Detailed experimental set-up for the odd/even test band for the comb-based silicon photonics data link. **b**, measured 100 Gbps PAM-4 eye diagrams from the sampling oscilloscope for each channel at receiving end.

Extended Data Fig. 4 Microcomb generation with different pump schemes.

a, Measured linewidth of the DFB laser. Comparison of the comb spectra pumped by **b**, a commercial external cavity laser and **c**, a DFB laser chip.

Extended Data Fig. 5 Dispersive delay-line MPF.

Experimental setup of the second dispersive delay-line scheme based TDL-MPF.

Extended Data Fig. 6 The response time measurement for the reconfiguration of RF filtering profiles.

a, Measured temporal response of the MRR under a square-wave electrical signal driving. **b**, MRR switched from minimum to maximum transmission, and **c**, maximum to minimum transmission. 90/10 rise/fall times are 15 μ s and 53 μ s, respectively.

Supplementary information

Supplementary Information

This Supplementary Information file contains Supplementary Sections 1–3, including Supplementary Figs. 1–3 and additional references. Section 1: Analysis of the dark pulse evolution. Section 2: Accessibility analysis under the thermal effects. Section 3: Theoretical fitting method for the RF filter responses of the MPF.

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

Shu, H., Chang, L., Tao, Y. *et al.* Microcomb-driven silicon photonic systems. *Nature* **605**, 457–463 (2022). <https://doi.org/10.1038/s41586-022-04579-3>

- Received: 04 August 2021
- Accepted: 24 February 2022
- Published: 18 May 2022
- Issue Date: 19 May 2022
- DOI: <https://doi.org/10.1038/s41586-022-04579-3>

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/s41586-022-04579-3>

[Download PDF](#)

- Article
- [Published: 18 May 2022](#)

Microfluidic chain reaction of structurally programmed capillary flow events

- [Mohamed Yafia^{1,2} na1](#),
- [Oriol Ymberri^{1,2} na1](#),
- [Ayokunle O. Olanrewaju^{1,2} nAff3](#),
- [Azim Parandakh^{1,2}](#),
- [Ahmad Sohrabi Kashani^{1,2}](#),
- [Johan Renault^{1,2}](#),
- [Zijie Jin ORCID: orcid.org/0000-0002-9579-1452^{1,2}](#),
- [Geunyong Kim^{1,2}](#),
- [Andy Ng ORCID: orcid.org/0000-0003-3257-0801^{1,2}](#) &
- [David Juncker ORCID: orcid.org/0000-0002-7313-1162^{1,2}](#)

[Nature](#) volume **605**, pages 464–469 (2022)

- 6993 Accesses
- 161 Altmetric
- [Metrics details](#)

Subjects

- [Biomedical engineering](#)
- [Lab-on-a-chip](#)

Abstract

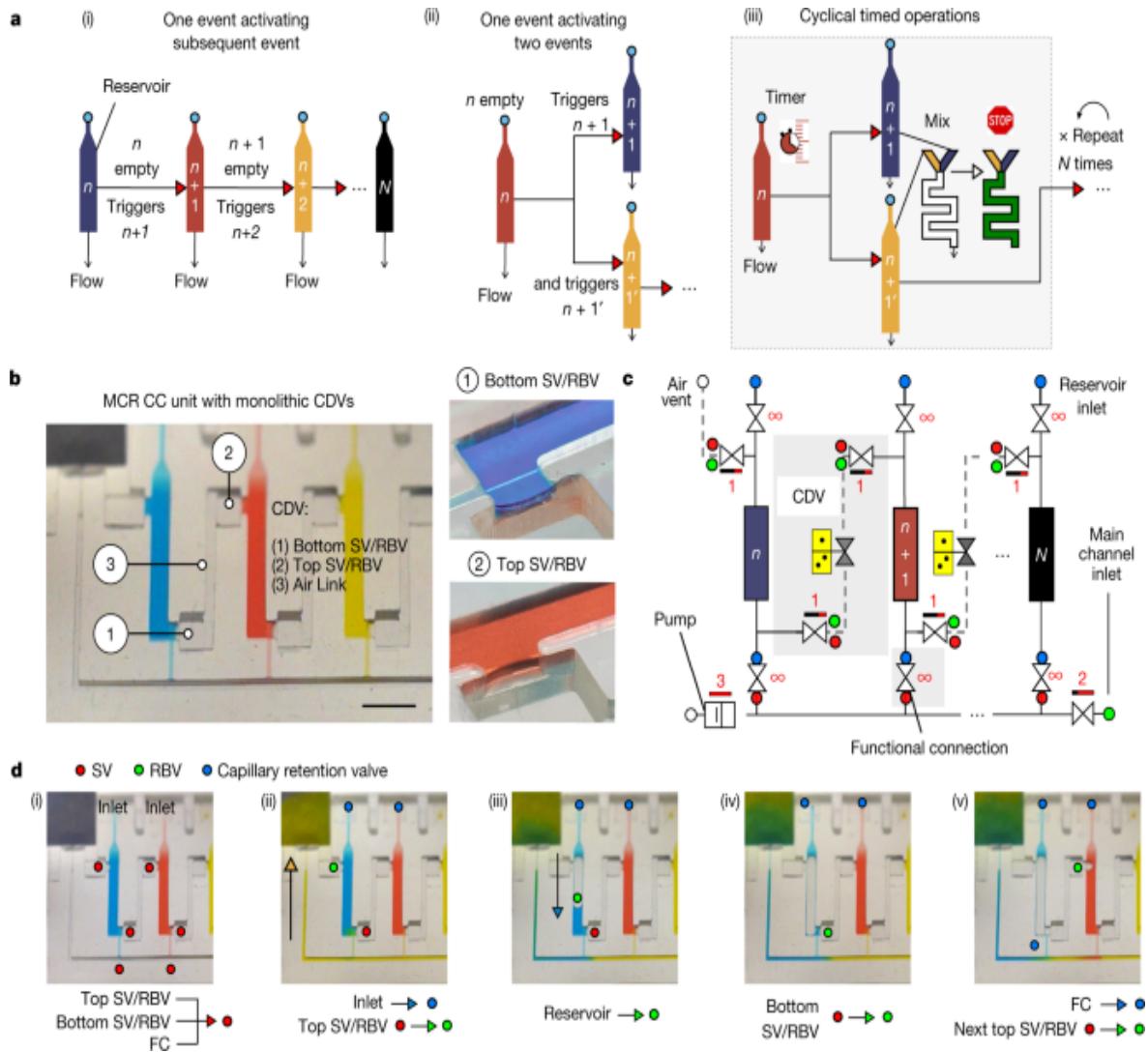
Chain reactions, characterized by initiation, propagation and termination, are stochastic at microscopic scales and underlie vital chemical (for example, combustion engines), nuclear and biotechnological (for example, polymerase chain reaction) applications^{1,2,3,4,5}. At macroscopic scales, chain reactions are deterministic and limited to applications for entertainment and art such as falling dominoes and Rube Goldberg machines. On the other hand, the microfluidic lab-on-a-chip (also called a micro-total analysis system)^{6,7} was visualized as an integrated chip, akin to microelectronic integrated circuits, yet in practice remains dependent on cumbersome peripherals, connections and a computer for automation^{8,9,10,11}. Capillary microfluidics integrate energy supply and flow control onto a single chip by using capillary phenomena, but programmability remains rudimentary with at most a handful (eight) operations possible^{12,13,14,15,16,17,18,19}. Here we introduce the microfluidic chain reaction (MCR) as the conditional, structurally programmed propagation of capillary flow events. Monolithic chips integrating a MCR are three-dimensionally printed, and powered by the free energy of a paper pump, autonomously execute liquid handling algorithms step-by-step. With MCR, we automated (1) the sequential release of 300 aliquots across chained, interconnected chips, (2) a protocol for severe acute respiratory syndrome-coronavirus-2 (SARS-CoV-2) antibodies detection in saliva and (3) a thrombin generation assay by continuous subsampling and analysis of coagulation-activated plasma with parallel operations including timers, iterative cycles of synchronous flow and stop-flow operations. MCRs are untethered from and unencumbered by peripherals, encode programs structurally in situ and can form a frugal, versatile, bona fide lab-on-a-chip with wide-ranging applications in liquid handling and point-of-care diagnostics.

Main

The MCR encodes the deterministic release of reagents stored in a series of reservoirs, with the release of reservoir n being conditional on the emptying (draining) of the reagent in reservoir $n - 1$, and emptying reservoir n , in turn triggering the release of reservoir $n + 1$. Capillary domino valves (CDVs)

encode this condition, and serially connect, that is, chain, the reservoirs, and thus control the propagation of the chain reaction (Fig. 1a). MCRs were implemented in three-dimensionally printed circuits made with a common stereolithography printer with feature size from 100 μm to 1.5 mm, hydrophilized using a plasma chamber (Extended Data Fig. 1 and 2), sealed with a plain cover and connected to a capillary pump made of paper (filter papers or absorbent pads). The paper was spontaneously wetted by aqueous solution drawn from the microfluidic circuit by releasing free energy stored in the paper surface, and this drove the chain reaction; expressed differently, the capillary pump generated a negative capillary pressure that was hydraulically transmitted back into the circuit through the main channel and serially drained side-reservoirs connected by a small conduit, called the functional connection (further described below). CDVs form air links between adjacent reservoirs, serially connecting them along a path parallel to the main channel, but interrupted by filled reservoirs that form liquid plugs between CDV air links. When the (first) reservoir connected to the air vent through a continuous air link is emptied, the plug is removed and the length of the air link propagates to the next filled reservoir in the MCR (Fig. 1a–d and Supplementary Video 1). This simple design structurally encodes the conditional propagation of capillary flow events and the step-by-step release of an arbitrary number N of reservoirs without peripheral connections or moving parts, and is further detailed in the [Supplementary Information](#).

Fig. 1: MCR with monolithic CDV for serial delivery of reagents in capillaric circuits (CCs).



a, (i) Serial MCR, (ii) branching MCR, (iii) cascaded, timed MCR. **b**, MCR unit with three reservoirs chained through CDVs and close-up of dual function SV/RBVs that keep liquid out of the CDV air link (forming a pneumatic connection) and prevent premature drainage. **c**, Symbolic view of the MCR unit with capillary retention valves (infinity symbol), CDV (grey overlay) that includes an air link, two SV/RBVs and functional connection. **d**, Screen shots of Supplementary Video 1 showing MCR sequences in which most of the capillary elements have dual functions, one during reagent loading, one during MCR propagation. (i) A loaded chip with liquids confined to the reservoirs by physical and capillary valves. (ii) MCR is triggered (the inlet becomes a capillary retention valve and the top SV becomes a RBV). (iii) Emptying of the first reservoir on bursting of the top RBV. (iv) The bottom SV momentarily becomes an RBV that bursts

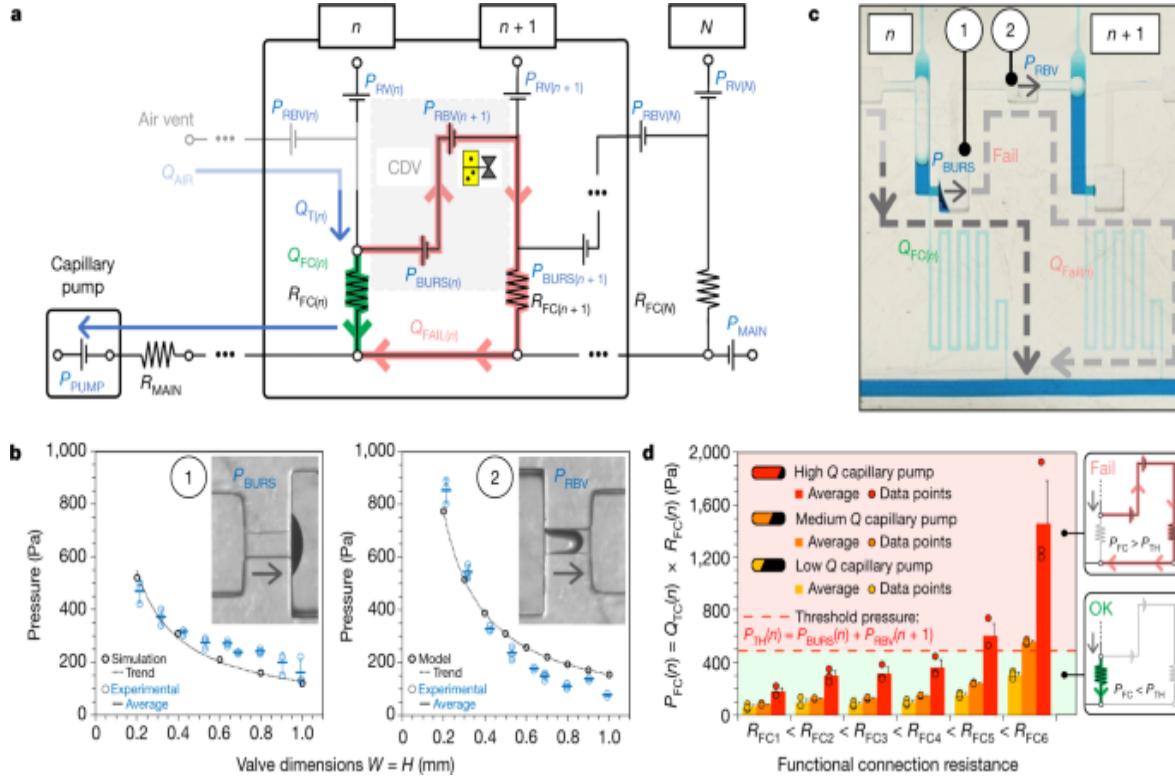
immediately. (v) Air now occupies the emptied reservoir. The functional connection (FC) becomes a capillary retention valve preventing the air from penetrating into the main channel. The air link connects the air to the RBV of the next reservoir, which bursts and triggers reservoir emptying. Scale bar, 2 mm.

MCRs require ancillary capillary microfluidic components that fulfil different functions depending on the intended operation (for example, loading, holding, mixing and draining liquids following the MCR progression) to form fully integrated and scalable capillaric circuits (CCs). CCs are designed on the basis of a library of building blocks including capillary pumps, flow resistances and many types of capillary valve (stop valves (SVs), trigger valves, retention valves, retention burst valves (RBVs))^{12,14}, and thus are analogous to microelectronic integrated circuits, but lacking the scalability and functionality. In MCRs, samples are loaded by capillary flow through an inlet with a capillary retention valve and entirely fill the reservoirs lined with three SVs, including two with a dual RBV function connecting to the two lateral CDVs, and one at the intersection of the functional connection and the main channel (Fig. 1c). Although the functional connection is a deceptively simple straight channel, it fulfils six key functions. It is (1) the air vent during filling of the reservoir, and (2) a SV preventing the reagent from spilling into the main channel while it is empty. After filling of the main channel, it forms a (3) hydraulic link propagating the pressure from the main channel into the reservoir and (4) a barrier (and bottleneck) to the diffusion of reagents between the reservoir and the main channel. (5) It becomes the outlet and a flow resistance (discussed further below) during reservoir emptying, and (6) a capillary retention valve stopping air from invading the main conduit after the reservoir is emptied. As a result, many trade-offs guide its design.

We sought to understand the design window and failure modes of MCRs, notably under which conditions downstream of CDVs might trigger prematurely, using both theory and experiments. MCR-CCs incorporate numerous capillary SVs according to previously established design criteria¹³ and while considering three-dimensional (3D) printer performance including resolution, imprecision and printing errors. We then analysed the MCR based on an electrical circuit analogy (Extended Data Fig. 3) and derived a

simplified circuit that neglects minor resistances (Fig. 2a)¹³. Successful and incremental propagation of the MCR is conditional on preventing the breach of the liquid in reservoir n into the CDV and air link connecting $n + 1$, which is equivalent to stating that all the liquid in reservoir n must flow exclusively through the functional connection n .

Fig. 2: Circuit analysis and experiments identify operational window for MCRs.



a, The simplified equivalent electrical circuit of the MCR units shown in Fig. 1. **b**, Experimental SV burst pressure (1) and RBV retention pressure (2) for valves with conduits with different, square cross-sections fitted with a numerical and an analytical model, respectively. **c**, Illustration of failure for a CDV with long serpentine FCs with very high resistance leading to liquid breach inside the air link, and premature draining of reservoir $n + 1$. **d**, Tests of six MCRs with increasing R_{FC} and three different paper pumps to determine the effect of varying the flow rate ($n = 3$ for each paper pump and R_{FC}). All data points are shown in **b** and **d**. Error bars are standard deviations from three experiments, the centre of each error bar is the mean

value. As predicted, the CDVs fail when the pressure drop across the FC $P_{FC(n)}$ exceeds the CDV threshold pressure $P_{BURS(n)} + P_{RBV(n+1)}$.

Source data

The flow path from reservoir n to $n + 1$ is interrupted by the CDV, which includes the capillary SV at one extremity and RBV at the other, with bursting thresholds of P_{BURS} and P_{RBV} , respectively. If either of these valves fails prematurely, then the propagation of the MCR is at risk of disruption. But because both valves are pneumatically connected by the air trapped within the air link, their pressures are additive and hence the threshold for failure of either is the sum of the two. The condition for success is $Q_{FAIL} = 0$, which during drainage of reservoir n is satisfied if the pressure drop on the functional connection (FC) $P_{FC} = Q_{FC} \times R_{FC}$ is (see also [Supplementary Information](#) for a detailed mathematical derivation):

$$\$ \$ \{ P \} _{\{ \{ \backslash rm{FC} \} \}(n)} < \{ P \} _{\{ \{ \backslash rm{BURS} \} \}(n)} + \{ P \} _{\{ \{ \backslash rm{RBV} \} \}(n+1)} \$ \$ \\ (1)$$

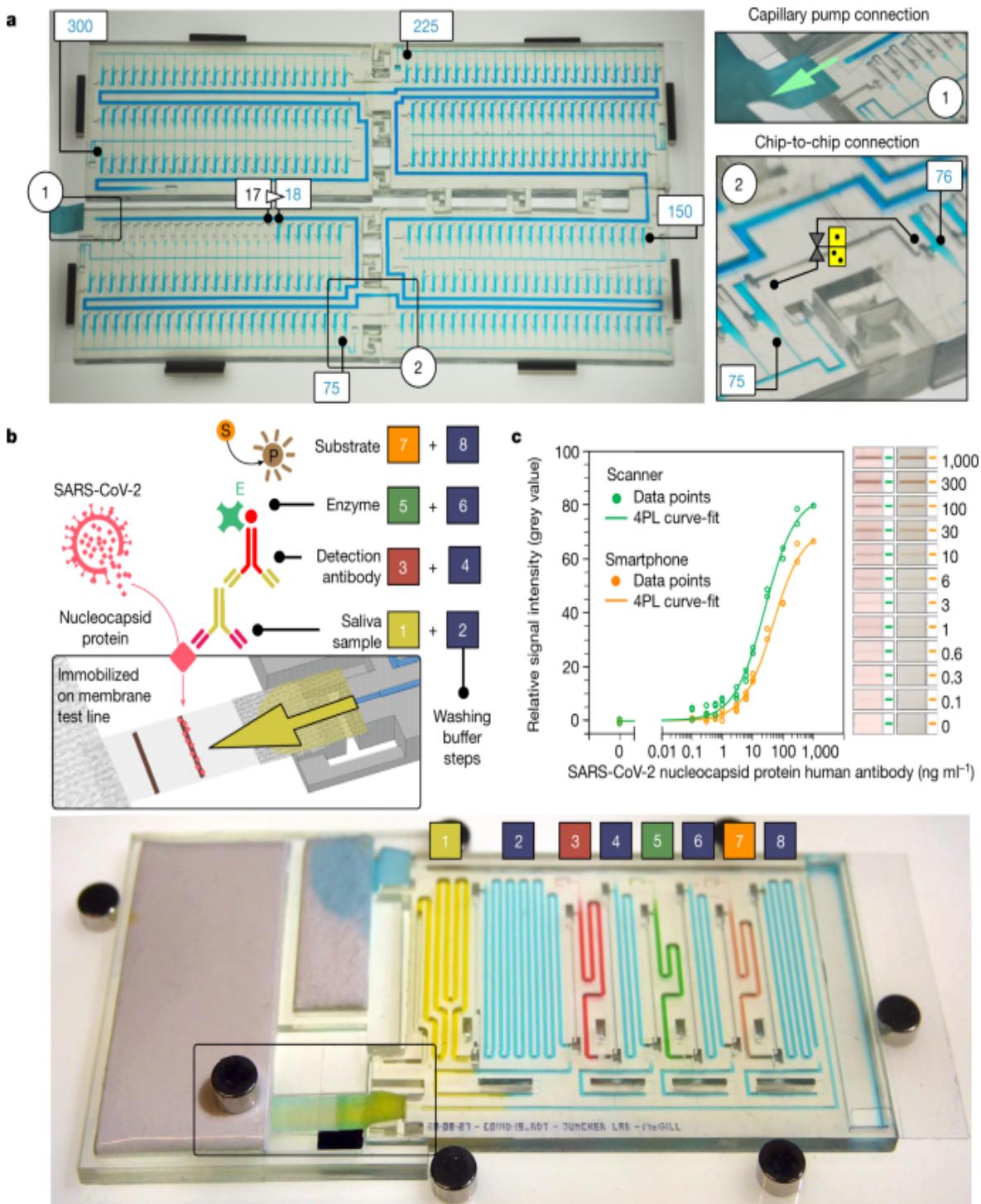
We calculated P_{BURS} (numerically)²⁰ and P_{RBV} (analytically, [Supplementary Information](#)) for conduits with a square cross-section ($W = H$) for the typical dimension in our 3D-printed CCs, and measured them experimentally for validation (Fig. 2b and Extended Data Fig. 4). Both P_{BURS} and P_{RBV} are inversely proportional to the smallest dimension of the rectangular conduit. We accounted for the hydrophobic ceiling formed by the sealing tape in both cases (Extended Data Fig. 2b), and which is a key feature to forming a functional SV²⁰. Note that because of the comparatively low pressures and small volume of the air links, the compressibility of air is negligible here.

Next, several MCRs featuring functional connections with large and increasing R_{FC} were tested with pumps with different capillary pressure and flow rates. The interplay between the resistance and the flow rate determines the operational window for the CDV while they are inversely proportional. We found excellent concordance between theory and experiments for the operation window of the MCR, and failure only occurred for the highest values of R_{FC} (nos. 5 and 6), and for only the most powerful capillary pumps

(Fig. [2c,d](#) and Extended Data Fig. [5](#)). The MCR designs used in the proof-of-concept applications, shown below, are well within the failure threshold, helping to ensure reliable propagation of the chain reaction.

We designed a chip-to-chip interface with a leakage-free connection for liquid (main channel) and air (connecting the CDVs), respectively, and connected four chips with 75 MCRs each (Fig. [3a](#) and Supplementary Video [2](#)). This result illustrates the reliability of the MCR and of CDVs, and demonstrates integrated, large-scale fluidic operations by ‘passive’ capillary microfluidics, beyond the capability of many ‘active’, computer programmable microfluidic systems.

Fig. 3: Large-scale MCR and COVID-19 serology assay in saliva.



a, A MCR of 300 aliquots stored in 4.9 μ l reservoirs across four chained and interconnected chips (Supplementary Video 2). **b**, SARS-CoV-2 antibody detection in saliva. Sequential, preprogrammed release of reagents by MCR is triggered by connecting the paper pump (Supplementary Video 3). The MCR supplies four reagents and four buffers in sequence. The functionality

includes delivery and removal (by flushing) of solutions, metering (40–200 μ l) by reservoir size, flow speed and time control by the flow resistance of the FC and the capillary pressure of the paper pump. The enzymatic amplification produces a brown precipitate line visible to the naked eye. **c**, Assay results and binding curve obtained by spiking antibody into saliva, and imaging by scanner and cell phone with representative images of the detection zone for each concentration, indicating the potential for quantitative point-of-care assays. **d**, An assembled chip filled with coloured solutions highlighting the channels for the different reagents and washing buffer.

[Source data](#)

Automated SARS-CoV-2-specific saliva antibody detection assay

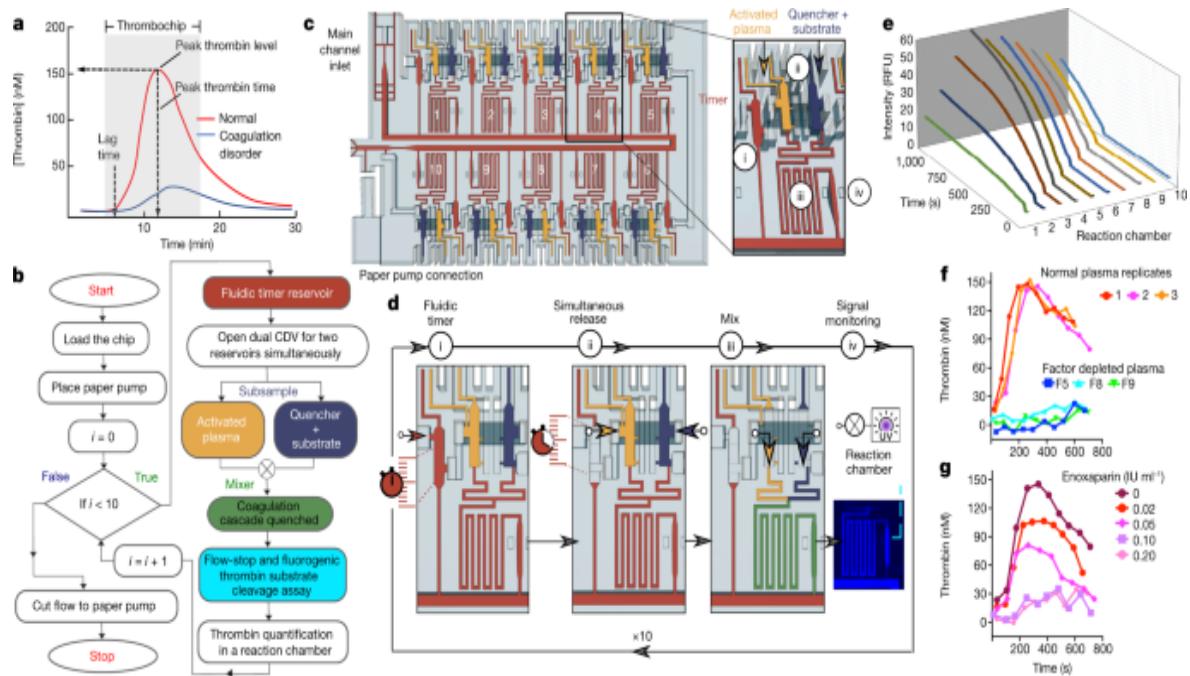
We measured antibodies against the nucleocapsid protein (N protein) of SARS-CoV-2 in saliva, with application potential for early infection detection^{21,22}, initial patient assessment as prognosis indicator²³ and for serosurveys to differentiate vaccinated and naturally infected individuals²⁴. Conventional lateral flow assays with predried reagents are simple to operate, but typically do not include enzymatic amplification that underlies the laboratory enzyme-linked immunosorbent assay (ELISA), and have to be read out within a few minutes of completion. Here, we used MCR to automate a sequence of eight steps in common laboratory ELISA protocols (Fig. [3b](#) and Supplementary Video [3](#)). The chip is connected to a small paper pump to drain excess buffer, and a nitrocellulose strip for assay readout itself connected to a large-capacity paper pump that drives the MCR. Note that the MCR propagates in a direction opposite from the flow in the main channel, and reagents released sequentially from reservoirs all flow past previously emptied reservoirs, thus minimizing the diffusional mixing between reagents. We used 3,3'-diaminobenzidine as a substrate that on enzymatic conversion produced a brown, persisting precipitate that could serve both as an immediate readout and a record for archival. Assay parameters such as volume, time and reagent concentrations were optimized extensively following standard protocols (see Extended Data Fig. [6](#) for examples) and

will be reported elsewhere. The result can be visualized by the naked eye or quantified using a scanner or a smartphone integrated with a simple folded origami box to minimize light interference, with a sensitive, quantitative and reproducible output (Fig. 3c and Extended Data Figs. 7 and 8).

Automated microfluidic thrombin generation assay (TGA)

Routine coagulation tests (prothrombin time and activated partial thromboplastin time) are used as initial evaluation of haemostatic status. These tests terminate on clot formation and thus only inform on the initiation of clotting, whereas the coagulation cascade continues and generates 95% of total thrombin (the final enzyme in the coagulation cascade)²⁵. The haemostatic capacity, expressed as the endogenous thrombin potential, can therefore not be fully evaluated by these tests²⁶. Global coagulation assays, such as the TGA that provides the time-course of active thrombin concentration in clotting plasma, are better measures of haemostatic function. Peak height, shape and area under curve of the thrombin generation curve (also known as the thrombogram, Fig. 4a) can be determined and correlated to clinical phenotypes to investigate coagulation disorders, and measure the effect of anticoagulants²⁷. The first TGA was introduced in the 1950s, and involves the activation of coagulation of blood or plasma, followed by a two-stage assay that requires the collection and mixing of subsamples with fibrinogen (or chromogenic substrates following their availability) at precisely timed intervals (for example, 1 min) over the course of 20 min or so, followed by the quantification of thrombin in each of them^{28,29}. The labour intensity, strict timing requirements and risk of error are great obstacles to wider adoption and clinical use of TGAs-by-subsampling. The calibrated automated thrombogram (CAT) introduced in 2002 simplifies operations thanks to newly synthesized thrombin substrates, a calibration TGA using the patient sample spiked with reference material and mathematical extrapolation³⁰.

Fig. 4: Automated TGA by continuous analysis of plasma subsamples (thrombochip).



Source data

Here, we demonstrate the capacity of MCR to automate the original TGA-by-subsampling in a microfluidic implementation that we called a "thrombochip". We devised an algorithm (Fig. 4b) for automating and timing the procedure with cascaded, iterative and branching fluidic

operations, and structurally programmed it into a 3D-printed chip (Fig. 4c, Extended Data Fig. 9 and Supplementary Video 4). Defibrinated, coagulation-activated plasma subsamples and reagent were loaded into the thrombochip, and on triggering of the MCR, without further intervention, they were released at 1 min intervals from the ten pairs of reservoirs, mixed in the serpentine mixer and stored in a 2.1 μ l reaction chamber with a width of 500 μ m for fluorescence signal generation and readout using a camera (Supplementary Videos 5 and 6). The concentration of thrombin in each of the subsamples is proportional to the rate of the fluorescent substrate turnover, and the time-course of thrombin is reported as a thrombogram.

Reliable execution of the TGA subsample analysis algorithm faced several practical challenges, and in particular draining of two reservoirs simultaneously is inherently unstable. Indeed, as soon as one reservoir starts being drained, the (absolute) pressure in the CC drops, and readily falls below the threshold of the RBV of the second reservoir, which will not burst, meaning the reservoir will remain filled. The MCR and 3D printing helped overcome this challenge and the reservoir pair containing plasma and reagents could be drained synchronously. An embedded air link connecting the outlet of reservoir n to the RBVs of both $n + 1$ and $n + 1'$, which were identical and very weak RBVs (cross-section, 1 \times 1 mm²) lead to simultaneous bursting and reliable propagation of the chain reaction. Other critical features are a serpentine mixer; stop-flow and holding of the solution in the reaction chambers for the thrombin quantification; a pressure pinning structure at the main outlet to cut the hydraulic connection to the paper pump after completion of the fluidic operation; an RBV at the main outlet that pins liquid and helps prevent backflow to safeguards the reaction chambers from uncontrolled mixing and finally evaporation during the extended monitoring and imaging of the thrombin reaction.

As validation of the thrombochip, human pooled plasma, plasma depleted of Factors V, VIII and IX, and plasma spiked with the anticoagulant Enoxaparin (an anti-Factor Xa drug) were analysed. The corresponding thrombograms were reproducible, consistent with normal and impaired coagulation cascades caused by factor depletion, and measured the dose-response of Enoxaparin (Fig. 4f,g). The general profile of the thrombograms generated in these proof-of-concept experiments are comparable to those by CAT and other microtitre plate-based assays^{31,32}, but direct comparison of

the data such as lag time and peak concentration requires standardized sample processing, reference materials and normalization, which can guide future development of the thrombochip.

Conclusion and discussion

MCRs introduce deterministic, modular and programmable chain reactions at the mesoscale and constitute a new concept for autonomous, programmable liquid operations and algorithms by control of both hydraulic and pneumatic flow and connectivity. The automation of complex and repetitive liquid handling operations has so far only been possible with a computer, software programs and cumbersome peripheral equipment, either robotics or, in the case of microfluidics⁶, systems to supply reagents, power or flow control^{8,9,10,11}. MCR introduces mesoscale chain reactions as a frugal, integrated, scalable and programmable process that power integrated labs-on-a-chip.

The MCR chip micro-architecture is simultaneously the circuit and the code of the chain reaction, is manufacturable with a variety of techniques and scalable along two distinctive paths: First, following microelectronics example and Moore's law, by shrinking and increasing the number of features per unit area and per unit volume (for example, by using 3D printing). Second, by expanding the overall size of CC-MCRs by interconnecting and chaining chips, and, inspired by trees that draw liquids more than 100 m in height, linking them to powerful capillary pumps³³. We anticipate numbers of steps far beyond the 300 shown here, and far more complex algorithms than the ones of the thrombochip.

MCRs are generalizable, compatible with positive pressure operations and could be interfaced with active microfluidics and robotic liquid handling systems. Spontaneous, capillary flow MCRs may be further improved too with permanently hydrophilic resins or coatings, liquid storage pouches and predried reagents³⁴, notably for point-of-care applications and any other uses. An end-user, by simply depositing a drop of solution at the inlet, could trigger a choreography of timed operations including aliquoting, delivery, mixing, flushing and reactions of several chemicals. As MCRs can be 3D-printed and monolithically encoded in a chip, the entry barrier is very low

(entry-level resin-based printers cost <US\$ 300). MCRs may be home-manufactured easily, or mail-ordered, opening the way for rapid dissemination and for new inventions, advances and for downloadable and printable microfluidic apps.

Methods

Chip design and fabrication

The chips were designed using AutoCAD (Autodesk) and exported as .STL files for 3D printing. CCs encoding MCRs were made with a digital micromirror display (DMD) 3D printer (Miicraft 100, Creative Cadworks) using a transparent resin (Rapid Model Resin Clear, Monocure 3D) purchased from filaments.ca. The following printing parameters were used: the layer thickness was 20 μm and the exposure time 1.5 s per layer, whereas the exposure time for the base layer was 10 s with four transition buffer layers. Following completion of the print, the chips were cleaned with isopropanol and post-cured for 1 min under ultraviolet (UV) light (Professional CureZone, Creative Cadworks).

Microchannels with cross-sections ranging from 250×100 to $1,500 \times 1,000 \mu\text{m}^2$ were fabricated and hydrophilized by plasma activation for 10 s at approximately 30% power (PE50 plasma chamber, Plasma Etch).

CCs were sealed with a delayed tack adhesive tape (9795R microfluidic tape, 3M) forming the cover.

Paper capillary pump

Filter papers (Whatman filter paper grade 4, 1 and 50 Hardened, Cytiva) were used as paper capillary pumps for all experiments except the SARS-CoV-2 antibody assay. The pore size from 4, 1 and 50 hardened is in decreasing order, and flow resistance and capillary pressure increase with decreasing pore size.

For the SARS-CoV-2 antibody assay, absorbent pads (Electrophoresis and Blotting Paper, Grade 238, Ahlstrom-Munksjo Chromatography) were used

as pumps.

Chip-to-chip connections for the 300 capillary flow events

To obtain a leakage-free connection, a thin layer of uncured photoresin, prepared by mixing poly (ethylene glycol) diacrylate (PEG-DA MW 258, Sigma-Aldrich) and Irgacure-819 (1% w/w), was applied to all of the chip-to-chip interfaces. Next, the chips were assembled and exposed to UV light in a UV chamber (320–390 nm, UVitron Intelliray 600) at 50% intensity for 30 s to cure the resin and seal the connections.

Videos and image processing

Videos and images were recorded using a Panasonic Lumix DMC-GH3K. Structural images of the chip and the embedded conduits were obtained using micro-computed tomography (Skyscan 1172, Bruker) and used to confirm the dimensions. Contact angles were measured on the basis of side view images ($n = 3$) and analysed using the Dropsnake extension in Image J.

Modelling and calculations

The theoretical burst pressures of capillary SVs were calculated by solving the flow field using the finite element method with COMSOL Multiphysics v.5.5. Experimentally measured contact angles (100° and 40° for the cover and the channel, respectively) were used to solve two-phase capillary flow using the level-set method. The capillary flows leading up to the SV was solved for a time period of 0–0.02 s with a time step of 1×10^{-5} s. The inlet pressure was varied with 10 Pa increment for each simulation until a burst was observed.

Experiments on pressure thresholds for capillary SV and RBV

We 3D-printed modules to evaluate SV/RBV with different cross-section areas. Each module contained three SV/RBV for replicate results. SV/RBV consisted of a two-level SV based on a geometrical channel expansion, as described elsewhere¹². The chips integrated a conical inlet/outlet for tubing connection to a microfluidic flow controller system (MFCS-4C) and

Fluiwell package (Fluigent) with fluidic reservoirs containing 5% red food dye in MilliQ water solution (see Extended Data Fig. 4 for setup images and Fig. 2 for contact angles). MAESFLO v.3.3.1 software (Fluigent) controlled the application of positive or negative pressure to calculate the burst pressures of the SV (liquid burst into air link) and RBV (receding meniscus), with increments of 0.1 mbar (roughly 10 Pa).

SARS-CoV-2 antibody assay

Reagents

SARS-CoV-2 nucleocapsid protein was purchased from Sino Biological, Inc. (40588-V08B). Human Chimeric antibody against SARS-CoV-2 nucleocapsid protein was purchased from Genscript Biotech (A02039). SIGMAFAST 3,3'-diaminobenzidine tablets were purchased from Sigma-Aldrich. Biotinylated Goat-anti-Human antibody was purchased from Cedarlane (GTXHU-003-DBIO). Pierce streptavidin poly-HRP (21140) was purchased from ThermoFisher.

Nitrocellulose strips

Nitrocellulose membranes (Whatman FF80HP Plus nitrocellulose-backed membranes, Cytiva) were cut into 5.2-mm-wide strips using the Silhouette Portrait paper cutter (Silhouette). Membranes were striped with a 5-mm-wide test line of 0.25 mg ml^{-1} SARS-CoV-2 nucleocapsid protein delivered using a programmable inkjet spotter (sciFLEXARRAYER SX, Scienion). The test line consists of four lanes of 50 droplets of about 350 pl printed 100 μm apart from each other. Eight passes of 25 droplets were used for each lane on even and odd positions to allow solution absorption in between passes. The membranes were then dried for 1 h at 37 °C before blocking by dipping into 1% BSA in 1× PBS solution until completely wet, then retrieved and left to dry for 1 h at 37 °C and then stored with desiccant at 4 °C until use the next day.

Connection of capillary pump and nitrocellulose chip to MCR chips

Nitrocellulose strips were mounted following standard lateral flow assay assembly protocols. The nitrocellulose strip was connected to a glass fibre conjugate pad (G041 SureWick, Millipore Sigma) on one end, and to an absorbent pad (Electrophoresis and Blotting Paper, Grade 238, Ahlstrom-Munksjo Chromatography) serving as the capillary pump at the other end. All three were attached to an adhesive tape serving as the backing layer. For the saliva antibody assay, the nitrocellulose strip was sandwiched between three absorbent pads ($15 \times 25 \text{ mm}^2$) and clamped with a paper clip. For the food-dye demonstrations a single absorbent pad ($25 \times 45 \text{ mm}^2$) was magnetically clamped to the nitrocellulose membrane.

Saliva assay protocol

Human saliva was extracted with oral swabs (SalivaBio, Salimetrics), followed by centrifugation and 1:10 dilution with $0.22 \mu\text{M}$ filtered phosphate buffer saline containing 1% BSA, 0.1% Tween 20. Human chimeric antibody against SARS-CoV-2 nucleocapsid protein at 0 to 1,000 ng ml^{-1} was spiked into diluted saliva and loaded to the sample reservoirs. Three replicate measurements for concentrations of 0–10 ng ml^{-1} , two replicate measurements for concentrations of 30–300 ng ml^{-1} and one measurement for 1,000 ng ml^{-1} . Biotinylated goat anti-human antibody at $0.5 \mu\text{g ml}^{-1}$ and streptavidin poly-HRP at $0.5 \mu\text{g ml}^{-1}$ were used to detect the human antibody. Control line in the nitrocellulose strip confirms reagents delivery and colorimetric reaction completion.

Image analysis on the nitrocellulose strips

After drainage of all reservoirs, the nitrocellulose membrane strip was removed, placed on a support and left to dry for 1 h.

The dry strips were imaged using (1) a flatbed scanner (mfc-9970cdw, Brother) at a resolution of 600 dpi and (2) using a Huawei P10 smartphone with a 12 megapixel image sensor and a rear camera with a 27 mm focal length (Huawei) in a customized box. The box was cut and folded with black cardboard paper to block ambient light when imaging with the smartphone. The box had two slots fitting the size of camera and nitrocellulose strip,

respectively, to ensure accurate alignment of the strip for readout. Images were taken with on-camera dual tone light-emitting diode flash at full power. Analysis of smartphone-taken and scanned images was done as follows.

Mean grey values of nitrocellulose test lines were extracted with ImageJ 1.48v (ImageJ, public domain software, W. Rasband, National Institutes of Health) within a 100×10 pixel rectangular area. Local background grey values were taken at 2.5 mm (0.1 inch) above each test line (following direction of the flow) for the same rectangular area, and subtracted from test line values. The normalized standard curve was then generated by subtracting negative control signal value (0 ng ml^{-1}) from all data points.

The limit of detection was calculated using the three-sigma criterion using a non-linear four-parameter logistic curve fit of the log-transformed data with OriginPro 8.5 SPR (OriginLab Corporation).

Automated microfluidic TGA (Thrombochip)

Citrated human plasma (P9523, lot number SLBX8880), fluorogenic thrombin substrate Z-GGR-AMC and Enoxaparin were purchased from Sigma-Aldrich; Batroxobin was from Prospec; Technothrombin TGA RC High reagent was from Diapharma; Human thrombin, non-patient plasma that were immuno-depleted of Factor V and Factor IX, and Factor VIII inactivated were from Haematologic Technologies; (4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid) (HEPES), and ethylenediaminetetraacetic acid (EDTA) and CaCl_2 were from Sigma-Aldrich.

The purchased pooled human plasma (collected in the United States in a Food and Drug Administration licensed centre site no. 268, as specified in the Certificate of Origin supplied by the manufacturer) was prepared by the manufacturer from whole blood collected by standard industry method using 4% trisodium citrate as an anticoagulant, pooled and then centrifuged. The resulting plasma was $0.45 \mu\text{m}$ filtered and lyophilized. Factor V- and Factor IX-depleted plasma were immune depleted; Factor VIII-depleted plasma was prepared by chemical depletion. The plasma preparations were assayed to ensure the activity of the remaining factors by the manufacturer.

Human plasma (pooled normal or factor depleted) were defibrinated by the addition of batroxobin (final concentration 0.6 BU ml⁻¹). The mixtures were incubated at room temperature for 20 min, followed by an extra incubation at 4 °C for 1 h. The mixtures were then centrifuged at 10,000g for 10 min to remove the fibrin clot and other debris. Defibrinated plasma were collected from the supernatant.

A solution containing 21% defibrinated plasma (plasma defibrination is needed to prevent clogging of the microfluidic channels by the fibrin clot), 48% Technothrombin TGA RC High reagent (high phospholipid and relipidated tissue factor content) and 20 mM CaCl₂ in 25 mM HEPES at pH 7.4 was loaded into the sample reservoirs of the thrombochip. A substrate solution containing 420 µM Z-GGR-AMC, 30 mM EDTA in 25 mM HEPES at pH 7.4 was loaded into the reagent reservoirs. The concentration of plasma, activation agent and substrate were optimized to yield a peak thrombin concentration and time of 150 nM and 200 s. All solutions were equilibrated to room temperature for 20 min before loading. Coagulation-inhibited plasma contained Enoxaparin at final concentrations of 0 to 1.0 anti-Xa units ml⁻¹ or IU ml⁻¹. The samples and reagents were loaded on the chip after initiating the coagulation cascade. The paper pump was connected to the chip to start the flow after 5 min from initiating the coagulation cascade. Fluorescence signals generated in the reaction chambers were monitored by illuminating the thrombochip with UV light at 365 nm with 20 W (realUV LED Flood Light, Waveform Lighting) and the visible 440 nm fluorescence emission signals measured by imaging at 5 s intervals using a Panasonic Lumix DMC-GH3K digital camera (f/3.5, Exposure time: 2 s, ISO-200). The rate of fluorescence signal generation in each reaction chamber (that is, the slope of the recorded fluorescence generation curve) is a measure of the rate of substrate turnover by thrombin and was used to deduce the amount of thrombin generated using a standard curve. Image J was used to analyse the images for fluorescence intensity.

Standard curve for thrombin quantification

Ten human thrombin solutions at concentrations ranging from 0 to 300 nM in 25 mM HEPES at pH 7.4 were loaded into the ten sample reservoirs in the thrombochip. A substrate solution containing 420 µM Z-GGR-AMC, 30 mM

EDTA in 25 mM HEPES at pH 7.4 was loaded into the reagent reservoirs. The standard curve was constructed by plotting the slope of the recorded fluorescence generation curve in each reaction chamber against the known thrombin concentration of the solution that was loaded to the corresponding sample reservoir.

Reporting summary

Further information on research design is available in the [Nature Research Reporting Summary](#) linked to this paper.

Data availability

The 3D design files of the MCR-CC chips are included as part of this article, and are also available for download along with more images and descriptions at

<https://www.thingiverse.com/junckerlab/collections/microfluidic-chain-reaction-of-structurally-programmed-capillary-flow-events>. Data not presented in the article or supplementary material will be available upon request. [Source data](#) are provided with this paper.

References

1. Denisov, E., Sarkisov, O. & Likhtenshtein, G. I. *Chemical Kinetics: Fundamentals and Recent Developments* (Elsevier, 2003).
2. *Award Ceremony Speech. Nobel Media AB 2021* (Nobel Prize Organization, 2021);
<https://www.nobelprize.org/prizes/chemistry/1956/ceremony-speech/>
3. Mullis, K. et al. Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction. *Cold Spring Harb. Symp. Quant. Biol.* **51**, 263–273 (1986).
4. Saiki, R. K. et al. Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science* **239**, 487–491 (1988).

5. Dirks, R. M. & Pierce, N. A. Triggered amplification by hybridization chain reaction. *Proc. Natl Acad. Sci. USA* **101**, 15275–15278 (2004).
6. Sackmann, E. K., Fulton, A. L. & Beebe, D. J. The present and future role of microfluidics in biomedical research. *Nature* **507**, 181–189 (2014).
7. Manz, A., Graber, N. & Widmer, H. M. Miniaturized total chemical analysis systems: a novel concept for chemical sensing. *Sens. Actuators B. Chem.* **1**, 244–248 (1990).
8. Thorsen, T., Maerkl, S. J. & Quake, S. R. Microfluidic large-scale integration. *Science* **298**, 580–584 (2002).
9. Choi, K., Ng, A. H., Fobel, R. & Wheeler, A. R. Digital microfluidics. *Annu. Rev. Anal. Chem.* **5**, 413–440 (2012).
10. Strohmeier, O. et al. Centrifugal microfluidic platforms: advanced unit operations and applications. *Chem. Soc. Rev.* **44**, 6187–6229 (2015).
11. Easley, C. J. et al. A fully integrated microfluidic genetic analysis system with sample-in-answer-out capability. *Proc. Natl Acad. Sci. USA* **103**, 19272–19277 (2006).
12. Olanrewaju, A., Beaugrand, M., Yafia, M. & Juncker, D. Capillary microfluidics in microchannels: from microfluidic networks to capillaric circuits. *Lab Chip* **18**, 2323–2347 (2018).
13. Olanrewaju, A. O., Robillard, A., Dagher, M. & Juncker, D. Autonomous microfluidic capillaric circuits replicated from 3D-printed molds. *Lab Chip* **16**, 3804–3814 (2016).
14. Safavieh, R. & Juncker, D. Capillarics: pre-programmed, self-powered microfluidic circuits built from capillary elements. *Lab Chip* **13**, 4180–4189 (2013).
15. Ozer, T., McMahon, C. & Henry, C. S. Advances in paper-based analytical devices. *Annu. Rev. Anal. Chem.* **13**, 85–109 (2020).

16. Wang, X., Hagen, J. A. & Papautsky, I. Paper pump for passive and programmable transport. *Biomicrofluidics* **7**, 14107 (2013).
17. Kokalj, T., Park, Y., Vencelj, M., Jenko, M. & Lee, L. P. Self-powered imbibing microfluidic pump by liquid encapsulation: SIMPLE. *Lab Chip* **14**, 4329–4333 (2014).
18. Dal Dosso, F. et al. Creasensor: SIMPLE technology for creatinine detection in plasma. *Anal. Chim. Acta* **1000**, 191–198 (2018).
19. Achille, C. et al. 3D printing of monolithic capillarity-driven microfluidic devices for diagnostics. *Adv. Mater.* **33**, e2008712 (2021).
20. Glière, A. & Delattre, C. Modeling and fabrication of capillary stop valves for planar microfluidic systems. *Sens. Actuator A Phys.* **130-131**, 601–608 (2006).
21. Burbelo, P. D. et al. Sensitivity in detection of antibodies to nucleocapsid and spike proteins of severe acute respiratory syndrome coronavirus 2 in patients with coronavirus disease 2019. *J Infect. Dis.* **222**, 206–213 (2020).
22. Pisanic, N. et al. COVID-19 serology at population scale: SARS-CoV-2-specific antibody responses in saliva. *J. Clin. Microbiol.* **59**, e02204–e02220 (2020).
23. Batra, M. et al. Role of IgG against N-protein of SARS-CoV2 in COVID19 clinical outcomes. *Sci Rep.* **11**, 3455 (2021).
24. Beck, E. J. et al. Differentiation of Individuals Previously Infected with and Vaccinated for SARS-CoV-2 in an Inner-City Emergency Department. *J. Clin. Microbiol.* **60**, e02390–e02421 (2022).
25. Tripodi, A. Thrombin generation assay and its application in the clinical laboratory. *Clin. Chem.* **62**, 699–707 (2016).
26. Brummel-Ziedins, K. E. & Wolberg, A. S. Global assays of hemostasis. *Curr. Opin. Hematol.* **21**, 395–403 (2014).

27. Binder, N. B. et al. Clinical use of thrombin generation assays. *J. Thromb. Haemost.* **19**, 2918–2929 (2021).
28. Macfarlane, R. G. & Biggs, R. A thrombin generation test; the application in haemophilia and thrombocytopenia. *J. Clin. Pathol.* **6**, 3–8 (1953).
29. Pitney, W. R. & Dacie, J. V. A simple method of studying the generation of thrombin in recalcified plasma; application in the investigation of haemophilia. *J. Clin. Pathol.* **6**, 9–14 (1953).
30. Hemker, H. C. et al. The calibrated automated thrombogram (CAT): a universal routine test for hyper- and hypocoagulability. *Pathophysiol. Haemost. Thromb.* **32**, 249–253 (2002).
31. Gerotziafas, G. T. et al. Comparison of the effect of fondaparinux and enoxaparin on thrombin generation during in-vitro clotting of whole blood and platelet-rich plasma. *Blood Coagul. Fibrinolysis* **15**, 149–156 (2004).
32. Zavyalova, E. & Kopylov, A. Exploring potential anticoagulant drug formulations using thrombin generation test. *Biochem. Biophys. Rep.* **5**, 111–119 (2016).
33. Wheeler, T. D. & Stroock, A. D. The transpiration of water at negative pressures in a synthetic tree. *Nature* **455**, 208–212 (2008).
34. Gokce, O., Castonguay, S., Temiz, Y., Gervais, T. & Delamarche, E. Self-coalescing flows in microfluidics for pulse-shaped delivery of reagents. *Nature* **574**, 228–232 (2019).

Acknowledgements

We acknowledge J. Lessard-Wajcer, J. Pimprikar, M. Mirbagheri, Y. Paschalidis and V. Karamzadeh for their assistance, Y. Morocz for photography and T. Gervais for comments. This work was supported by NSERC Strategic Project grant no. STPGP 494495-16, NSERC Alliance grant no. ALLRP 551058-20 and an McGill MI4 SCRF grant. M.Y.

acknowledges FRQNT postdoctoral fellowship no. 260284. A.S.K.
acknowledges FRQNT postdoctoral fellowship no. 267919. D.J.
acknowledges support from a Canada Research Chair in Bioengineering.

Author information

Author notes

1. Ayokunle O. Olanrewaju

Present address: Mechanical Engineering Department, University of Washington, Seattle, WA, USA

2. These authors contributed equally: Mohamed Yafia, Oriol Ymbern

Affiliations

1. Biomedical Engineering Department, McGill University, Montreal, Quebec, Canada

Mohamed Yafia, Oriol Ymbern, Ayokunle O. Olanrewaju, Azim Parandakh, Ahmad Sohrabi Kashani, Johan Renault, Zijie Jin, Geunyong Kim, Andy Ng & David Juncker

2. McGill Genome Centre, McGill University, Montreal, Quebec, Canada

Mohamed Yafia, Oriol Ymbern, Ayokunle O. Olanrewaju, Azim Parandakh, Ahmad Sohrabi Kashani, Johan Renault, Zijie Jin, Geunyong Kim, Andy Ng & David Juncker

Contributions

Methodology, data curation and formal analysis were carried out by the following authors. D.J., A.O. and M.Y. designed and tested MCR-CCs. O.Y., A.P., J.R., Z.J., A.S. and A.P. designed MCR-CCs for experiments and analysed the data for the COVID-19 antibody test. O.Y. performed the pressure characterization tests. M.Y. and A.N. performed the experiments

and analysed the data for the thrombochip. M.Y., A.P. and A.S.K. designed and performed the experiments for the CDV failure. A.P. and A.S.K. performed the experiments for the 300 MCRs. M.Y. performed the computed tomography scans and dimensions characterization. A.S. and A.P. performed the contact angle measurements. G.K. performed the simulation for the burst pressure on COMSOL. D.J. guided design and analysis of all experiments. Visualization was carried out by M.Y, O.Y, A.N. and D.J. who prepared the main figures. M.Y. and Z.J. worked on the 3D visualizations. M.Y., O.Y., A.O., A.P., A.S.K., Z.J. and A.N. captured all the images and videos. G.K. and M.Y. edited the videos. The original draft was written by M.Y., A.O., A.N. and D.J. Review and editing of the draft were carried out by M.Y., O.Y., A.O., A.P., Z.J., J.R., G.K. A.N. and D.J. Project conception, administration and supervision was the responsibility of A.N. and D.J.

Corresponding author

Correspondence to [David Juncker](#).

Ethics declarations

Competing interests

M.Y., A.O. and D.J. are inventors on a report of invention submitted to McGill and licensed. All other authors declare no competing interests.

Peer review

Peer review information

Nature thanks Mazher Iqbal Mohammed, Wouter van der Wijngaart, Kevin Ariën 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 Designed and measured dimensions of 3D-printed channels.

(a) CT scan channels with square cross-section with different width and depth. (b) Measured depth of 3D-printed channels matches the nominal, designed depth. (c) The measured width of 3D-printed channels is $\sim 40 \mu\text{m}$ smaller than the designed width. Dashed line shows the 1:1 ratio. All data points are shown. Error bars (presented on the right side of the data points for clarity) are standard deviation of the measurement of 3 replicate channels; the centre of each error bar is the mean measurement value.

[Source data](#)

Extended Data Fig. 2 Contact angles of water without and with food dyes on materials used for MCR CC.

(a) Contact angle of Milli Q water and different food dyes used in the experiments on the 3D-printed resin used for making the MCR CCs that were pristine (without plasma treatment) and following plasma treatment for 10 s at 30% and 100% plasma power. (b) Contact angle on the hydrophobic tape (3M microfluidic tape 9795R) used as cover. All data points are shown. Error bars are standard deviation of 6 replicate measurements; the centre of each error bar is the mean value.

[Source data](#)

Extended Data Fig. 3 Equivalent resistance and pressure source diagram for an MCR unit including capillary domino valve (CDV) and functional connection (FC).

The pressure drop $P_{FC} = Q_{T(n)} \times R_{FC}$ with $Q_{T(n)}$ the total flow rate which is equal to Q_{FC} if all the flow flows through the functional connection, and also equal to Q_{Air} drawn in to replace the liquid drained from the reservoir, and is the condition for successful operation. The overall resistance of the main channel and capillary pump is R_{MAIN} . In our designs, the resistances for the reservoir R_{RES} , the stop valve R_{SV} , the retention burst valve R_{RBV} and the sections of the main channel $R_{CH(n)}$ are much smaller than R_{MAIN} and R_{FC} , and hence can be neglected when calculating $Q_{T(n)} \approx P_{PUMP}/(R_{MAIN} + R_{FC})$ and are also negligible for the CDV failure analysis and ignored in the simplified schematic of Fig. 2a. The capillary retention valve threshold pressure is indicated as P_{RV} for side channels and P_{MAIN} for the main channel, and their pressure will match the one in the liquid (relative to air) up the bursting threshold of the valve and are designed such that they do not burst. The retention burst valve pressure is P_{RBV} and the stop valve pressure is P_{BURS} and both will adopt the pressure value difference between the air link and the adjacent liquid, up to their bursting threshold. Failure of either means failure of the other as they are pneumatically linked (until reservoir n is emptied, and the condition for chain reaction propagation is met, that is), and also failure of the CDV and the MCR because liquid will penetrate into the CDV and the air link, and clog it, which is indicated by Q_{FAIL} . Hence the condition for successful operation and propagation of the MCR to the final reservoir N is $Q_{FAIL} = 0$ for each of the chain reaction steps.

Extended Data Fig. 4 SV/RBV burst pressure experiments and simulation.

(a) Experimental setup for the positive and negative burst pressure experiments for the stop valve (SV) and retention burst valve (RBV) respectively. We connected a microfluidic flow controller system to the 3D-printed microfluidic chip containing three geometric two-level stop valves, and different chips were fabricated with varied cross-section valves dimensions. We used a camera to record videos to evaluate the valve burst pressure tested with increments of 0.1 mbar (~ 10 Pa). **(b)** Close-up picture of a chip with the middle valve under evaluation. **(c)** Valve positive (left) and negative (right) burst pressure experiments. **(d)** Screenshots of the burst

pressure by finite element method simulation for the stop valve showing successful stop at 300 Pa, and bursting for 310 Pa.

Extended Data Fig. 5 Experimental operating window for CDV and MCR.

(a) Test chip for evaluating determining the operational window of the CDV and breach. The chip includes 6 functional connections with serpentines with constant width and decreasing channel depth of 500, 300, 240, 200, 140, and 100 μm , respectively, and hence increasing R_{FC} . The rightmost reservoir is needed to test the functional connection no. 6. **(b)** R_{FC} for each of the design variations ($R_{\text{FC}1}$ to $R_{\text{FC}6}$) (see also Fig. 2d). **(c)** The flow rate for each reservoir with a medium rate paper pump (Whatman filter paper 1). **(d)** With Whatman paper 1 pump, bursting of liquid into the CDV and air link was only observed for the last reservoir. **(e)** flow rate measurements for $R_{\text{FC}1}$ for the three different Whatman paper pumps with grade 4, 1 and hardened 50. All data points are shown in (c–e). Error bars are standard deviation of 3 replicate measurements, the centre of each error bar is the mean value. The capillary pump with the lowest flow resistance resulted in the highest flow rate. The CDV failure occurred when the pressure drop across the functional connection exceeds the threshold pressure (Fig. 2a).

[Source data](#)

Extended Data Fig. 6 SARS-CoV-2 antibody assay optimization examples.

(a) Reduction of background signal. The graph and nitrocellulose strips (below) show the grey value of the test line (TL) after running four chips with different reagent 3 (see Fig. 3b; biotinylated goat-anti-human antibody used as detection antibody) and reagent 5 (streptavidin-Poly-HRP as enzyme) concentrations. **(b)** Assay performance using non-optimized (red) and optimized concentrations of reagents 3 and 5 (green). Non-optimized conditions show high background and false positive signal on the test line for the negative control membrane strip.

[Source data](#)

Extended Data Fig. 7 SARS-CoV-2 antibody assay on nitrocellulose membranes.

Array of the nitrocellulose membrane strips images of the optimized conditions for the detection of SARS-CoV-2 nucleocapsid antibody in saliva at different concentrations, presented in Fig. [3c](#). On the left (**a**) pictures taken by a scanner (highlighted green test line) and right (**b**) by a smartphone camera (orange test line).

Extended Data Fig. 8 Foldable box for reducing light interference during smartphone-based test readout.

(**a**) Smartphone and folded cardboard setup used to measure the signal of the SARS-CoV-2 antibody assay. (**b**) Detail of smartphone picture taken using the foldable box. (**c**) Unfolded black box casing showing opening for smartphone camera lens and flash. (**d**) Assembled black box casing designed for images taken using smartphone with the built-in flash to analyze the signal line on the nitrocellulose membrane.

Extended Data Fig. 9 Micro computer tomography images of 3D printed thrombochip.

(**a**) micro CT scan at the cross-section of the reservoir and the embedded air link. (**b**) 3D schematic for one stage in the thrombochip design. (**c**) micro CT scan at the cross-section of the reaction chamber showing. (**d**) The fluorescence signal generated along the reaction chamber. The signal is measured at the deepest section, depicted as the measuring zone, to get the highest possible signal.

[Source data](#)

Supplementary information

Supplementary Information

MCR operational details; threshold pressure for capillary retention valves and RBVs; calculation of the failure condition for a CDV and of the MCR and legends for Supplementary Videos.

Reporting Summary

Peer Review File

Supplementary Video 1.

Supplementary Video 2.

Supplementary Video 3.

Supplementary Video 4.

Supplementary Video 5.

Supplementary Video 6.

Source data

Source Data Fig. 2

Source Data Fig. 3

Source Data Fig. 4

Source Data Extended Data Fig. 1

Source Data Extended Data Fig. 2

Source Data Extended Data Fig. 5

Source Data Extended Data Fig. 6

Source Data Extended Data Fig. 9

Rights and permissions

Reprints and Permissions

About this article

Cite this article

Yafia, M., Ymbern, O., Olanrewaju, A.O. *et al.* Microfluidic chain reaction of structurally programmed capillary flow events. *Nature* **605**, 464–469 (2022). <https://doi.org/10.1038/s41586-022-04683-4>

- Received: 01 June 2021
- Accepted: 24 March 2022
- Published: 18 May 2022
- Issue Date: 19 May 2022
- DOI: <https://doi.org/10.1038/s41586-022-04683-4>

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

[**A lab-on-a-chip that takes the chip out of the lab**](#)

- Mazher Iqbal Mohammed

Nature News & Views 18 May 2022

This article was downloaded by **calibre** from <https://www.nature.com/articles/s41586-022-04683-4>

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

[Download PDF](#)

- Article
- Open Access
- [Published: 18 May 2022](#)

Synergistic HNO₃–H₂SO₄–NH₃ upper tropospheric particle formation

- [Mingyi Wang](#) ORCID: orcid.org/0000-0001-5782-2513^{1,2 nAff7},
- [Mao Xiao](#)³,
- [Barbara Bertozzi](#) ORCID: orcid.org/0000-0003-2571-6585⁴,
- [Guillaume Marie](#) ORCID: orcid.org/0000-0003-1454-0908⁵,
- [Birte Rörup](#)⁶,
- [Benjamin Schulze](#)⁷,
- [Roman Bardakov](#)^{8,9},
- [Xu-Cheng He](#) ORCID: orcid.org/0000-0002-7416-306X⁶,
- [Jiali Shen](#)⁶,
- [Wiebke Scholz](#)¹⁰,
- [Ruby Marten](#) ORCID: orcid.org/0000-0003-0417-4350³,
- [Lubna Dada](#)^{3,6},
- [Rima Baalbaki](#) ORCID: orcid.org/0000-0002-4480-2107⁶,
- [Brandon Lopez](#)^{1,11},
- [Houssni Lamkaddam](#)³,
- [Hanna E. Manninen](#)¹²,
- [António Amorim](#)¹³,
- [Farnoush Ataei](#)¹⁴,
- [Pia Bogert](#)⁴,
- [Zoé Brasseur](#) ORCID: orcid.org/0000-0001-5387-018X⁶,
- [Lucía Caudillo](#)⁵,

- [Louis-Philippe De Menezes](#)¹²,
- [Jonathan Duplissy](#) ORCID: orcid.org/0000-0001-8819-0264^{6,15},
- [Annica M. L. Ekman](#) ORCID: orcid.org/0000-0002-5940-2114^{8,9},
- [Henning Finkenzeller](#)¹⁶,
- [Loïc Gonzalez Carracedo](#)¹⁷,
- [Manuel Granzin](#)⁵,
- [Roberto Guida](#)¹²,
- [Martin Heinritzi](#)⁵,
- [Victoria Hofbauer](#)^{1,2},
- [Kristina Höhler](#) ORCID: orcid.org/0000-0001-5503-9816⁴,
- [Kimmo Korhonen](#)¹⁸,
- [Jordan E. Krechmer](#) ORCID: orcid.org/0000-0003-3642-0659¹⁹,
- [Andreas Kürten](#)⁵,
- [Katrianne Lehtipalo](#) ORCID: orcid.org/0000-0002-1660-2706^{6,20},
- [Naser G. A. Mahfouz](#) ORCID: orcid.org/0000-0002-7097-1430^{1,21},
- [Vladimir Makhmutov](#)^{22,23},
- [Dario Massabò](#)²⁴,
- [Serge Mathot](#)¹²,
- [Roy L. Mauldin](#)^{1,2,25},
- [Bernhard Mentler](#) ORCID: orcid.org/0000-0003-3852-0397¹⁰,
- [Tatjana Müller](#) ORCID: orcid.org/0000-0001-6453-9134^{5,26},
- [Antti Onnela](#)¹²,
- [Tuukka Petäjä](#) ORCID: orcid.org/0000-0002-1881-9044⁶,
- [Maxim Philippov](#) ORCID: orcid.org/0000-0003-4302-0020²²,
- [Ana A. Piedehierro](#)²⁰,
- [Andrea Pozzer](#) ORCID: orcid.org/0000-0003-2440-6104²⁶,
- [Ananth Ranjithkumar](#)²⁷,
- [Meredith Schervish](#)^{1,2},
- [Siegfried Schobesberger](#)¹⁸,
- [Mario Simon](#)⁵,
- [Yuri Stozhkov](#)²²,
- [António Tomé](#)²⁸,
- [Nsikanabasi Silas Umo](#)⁴,
- [Franziska Vogel](#)⁴,

- [Robert Wagner⁴](#),
- [Dongyu S. Wang³](#),
- [Stefan K. Weber](#) [ORCID: orcid.org/0000-0001-7408-9069¹²](#),
- [André Welti](#) [ORCID: orcid.org/0000-0002-3549-1212²⁰](#),
- [Yusheng Wu⁶](#),
- [Marcel Zauner-Wieczorek](#) [ORCID: orcid.org/0000-0002-0867-665X⁵](#),
- [Mikko Sipilä⁶](#),
- [Paul M. Winkler¹⁷](#),
- [Armin Hansel](#) [ORCID: orcid.org/0000-0002-1062-2394^{10,29}](#),
- [Urs Baltensperger³](#),
- [Markku Kulmala](#) [ORCID: orcid.org/0000-0003-3464-7825^{6,15,30,31}](#),
- [Richard C. Flagan](#) [ORCID: orcid.org/0000-0001-5690-770X⁷](#),
- [Joachim Curtius](#) [ORCID: orcid.org/0000-0003-3153-4630⁵](#),
- [Ilona Riipinen^{9,32}](#),
- [Hamish Gordon](#) [ORCID: orcid.org/0000-0002-1822-3224^{1,11}](#),
- [Jos Lelieveld](#) [ORCID: orcid.org/0000-0001-6307-3846^{26,33}](#),
- [Imad El-Haddad](#) [ORCID: orcid.org/0000-0002-2461-7238³](#),
- [Rainer Volkamer](#) [ORCID: orcid.org/0000-0002-0899-1369¹⁶](#),
- [Douglas R. Worsnop^{6,19}](#),
- [Theodoros Christoudias](#) [ORCID: orcid.org/0000-0001-9050-3880³³](#),
- [Jasper Kirkby](#) [ORCID: orcid.org/0000-0003-2341-9069^{5,12}](#),
- [Ottmar Möhler⁴](#) &
- [Neil M. Donahue](#) [ORCID: orcid.org/0000-0003-3054-2364^{1,2,11,34}](#)

Nature volume **605**, pages 483–489 (2022)

- 585 Accesses
- 153 Altmetric
- [Metrics details](#)

Subjects

- [Atmospheric chemistry](#)
- [Atmospheric science](#)

- [Climate change](#)

Abstract

New particle formation in the upper free troposphere is a major global source of cloud condensation nuclei (CCN)^{[1,2,3,4](#)}. However, the precursor vapours that drive the process are not well understood. With experiments performed under upper tropospheric conditions in the CERN CLOUD chamber, we show that nitric acid, sulfuric acid and ammonia form particles synergistically, at rates that are orders of magnitude faster than those from any two of the three components. The importance of this mechanism depends on the availability of ammonia, which was previously thought to be efficiently scavenged by cloud droplets during convection. However, surprisingly high concentrations of ammonia and ammonium nitrate have recently been observed in the upper troposphere over the Asian monsoon region^{[5,6](#)}. Once particles have formed, co-condensation of ammonia and abundant nitric acid alone is sufficient to drive rapid growth to CCN sizes with only trace sulfate. Moreover, our measurements show that these CCN are also highly efficient ice nucleating particles—comparable to desert dust. Our model simulations confirm that ammonia is efficiently convected aloft during the Asian monsoon, driving rapid, multi-acid $\text{HNO}_3\text{--H}_2\text{SO}_4\text{--NH}_3$ nucleation in the upper troposphere and producing ice nucleating particles that spread across the mid-latitude Northern Hemisphere.

Main

Intense particle formation has been observed by airborne measurements as a persistent, global-scale band in the upper troposphere over tropical convective regions^{[1,2,4](#)}. Upper tropospheric nucleation is thought to provide at least one-third of global CCN^{[3](#)}. Increased aerosols since the industrial revolution, and their interactions with clouds, have masked a large fraction of the global radiative forcing by greenhouse gases. Projections of aerosol radiative forcing resulting from future reductions of air pollution are highly uncertain^{[7](#)}. Present-day nucleation involves sulfuric acid (H_2SO_4) over almost all the troposphere^{[8](#)}. However, binary nucleation of $\text{H}_2\text{SO}_4\text{--H}_2\text{O}$ is

slow and, so, ternary or multicomponent nucleation with extra vapours such as ammonia (NH_3)⁹ and organics^{10,11} is necessary to account for observed new-particle-formation rates^{3,8,12}.

Ammonia stabilizes acid–base nucleation and strongly enhances particle formation rates⁹. However, ammonia is thought to be extremely scarce in the upper troposphere because its solubility in water and reactivity with acids should lead to efficient removal in convective clouds. However, this assumption is not supported by observation. Ammonia vapour has been repeatedly detected in the Asian monsoon upper troposphere, with mixing ratios of up to 30 pptv ($2.5 \times 10^8 \text{ cm}^{-3}$) for a three-month average⁵ and up to 1.4 ppbv ($1.2 \times 10^{10} \text{ cm}^{-3}$) in hotspots⁶. The release of dissolved ammonia from cloud droplets may occur during glaciation¹³. Once released in the upper troposphere, ammonia can form particles with nitric acid, which is abundantly produced by lightning^{14,15}. These particles will live longer and travel farther than ammonia vapour, with the potential to influence the entire upper troposphere and lower stratosphere of the Northern Hemisphere⁶.

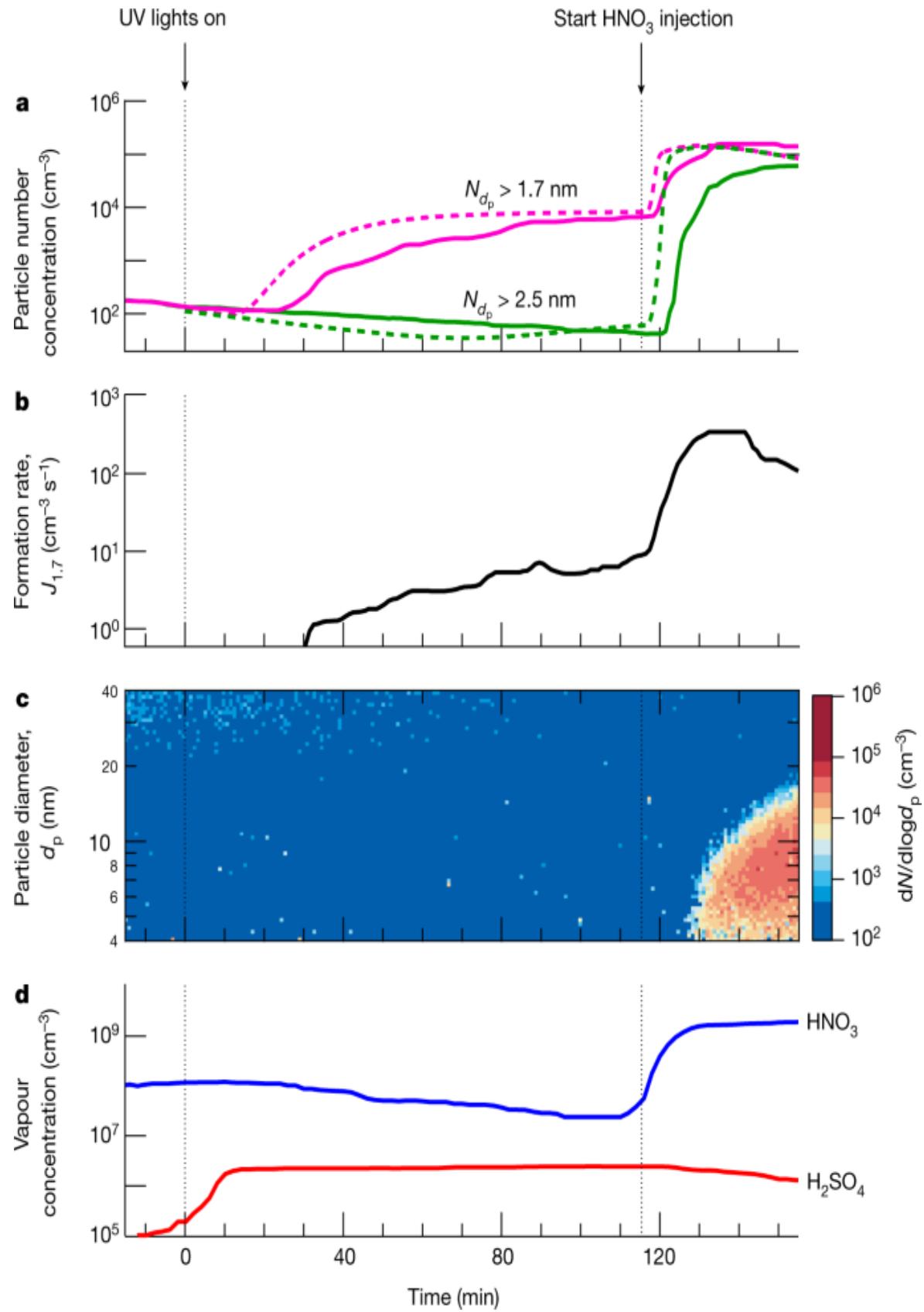
Fundamental questions remain about the role and mechanisms of nitric acid and ammonia in upper tropospheric particle formation. Recent CLOUD (Cosmics Leaving Outdoor Droplets) experiments at CERN have shown that nitric acid and ammonia vapours below 278 K can condense onto newly formed particles as small as a few nanometres in diameter, driving rapid growth to CCN sizes¹⁶. At even lower temperatures (below 258 K), nitric acid and ammonia can directly nucleate to form ammonium nitrate particles, although pure $\text{HNO}_3\text{--NH}_3$ nucleation is too slow to compete with $\text{H}_2\text{SO}_4\text{--NH}_3$ nucleation under comparable conditions. However, the results we present here show that, when all three vapours are present, a synergistic interaction drives nucleation rates orders of magnitude faster than those from any two of the three components. Once nucleated through this multi-acid–ammonia mechanism, the particles can grow rapidly by co-condensation of NH_3 and HNO_3 alone, both of which may be far more abundant than H_2SO_4 in the upper troposphere.

Particle formation measurements in CLOUD

Here we report new-particle-formation experiments performed with mixtures of sulfuric acid, nitric acid and ammonia vapours in the CLOUD chamber⁹ at CERN between September and December 2019 (CLOUD 14; see [Methods](#) for experimental details). To span ranges typical of the upper troposphere, we established quasi-steady-state vapour concentrations in the chamber of $(0.26\text{--}4.6) \times 10^6 \text{ cm}^{-3}$ sulfuric acid (through photochemical oxidation of SO_2), $(0.23\text{--}4.0) \times 10^9 \text{ cm}^{-3}$ nitric acid (through either photochemical oxidation of NO_2 or injection from an evaporator) and $(0.95\text{--}6.5) \times 10^8 \text{ cm}^{-3}$ ammonia (through injection from a gas bottle). In an extreme experiment to simulate hotspot conditions in the Asian monsoon anticyclone, we raised sulfuric acid, nitric acid and ammonia to maximum concentrations of $6.2 \times 10^7 \text{ cm}^{-3}$, $3.8 \times 10^9 \text{ cm}^{-3}$ and $8.8 \times 10^9 \text{ cm}^{-3}$, respectively. The experiments were conducted at 223 K and 25% relative humidity, representative of upper tropospheric conditions.

Figure 1 shows the evolution of a representative new-particle-formation experiment in the presence of around $6.5 \times 10^8 \text{ cm}^{-3}$ ammonia. The top three panels show particle number concentrations above 1.7 nm and above 2.5 nm (Fig. 1a), particle formation rate at 1.7 nm ($J_{1.7}$) (Fig. 1b) and particle size distribution (Fig. 1c). The bottom panel shows HNO_3 and H_2SO_4 vapour concentrations (Fig. 1d). We switched on the ultraviolet (UV) lights at $t = 0$ min to oxidize SO_2 with OH radicals and form H_2SO_4 . Sulfuric acid started to appear shortly thereafter and built up to a steady state of $2.3 \times 10^6 \text{ cm}^{-3}$ over the wall-loss timescale of about 10 min. Under these conditions, the data show a modest formation rate of 1.7-nm particles from $\text{H}_2\text{SO}_4\text{--NH}_3$ nucleation, consistent with previous CLOUD measurements⁸. These particles grew only slowly (about 0.5 nm h^{-1} at this H_2SO_4 and particle size¹⁷). No particles reached 2.5 nm within 2 h, owing to their slow growth rate and low survival probability against wall loss.

Fig. 1: Example experiment showing nitric acid enhancement of $\text{H}_2\text{SO}_4\text{--NH}_3$ particle formation.



a, Particle number concentrations versus time at mobility diameters >1.7 nm (magenta) and >2.5 nm (green). The solid magenta trace is measured by a PSM_{1.7} and the solid green trace is measured by a CPC_{2.5}. The fixed experimental conditions are about $6.5 \times 10^8 \text{ cm}^{-3}$ NH₃, 223 K and 25% relative humidity. A microphysical model reproduces the main features of the observed particle formation (dashed lines; see text for details). **b**, Particle formation rate versus time at 1.7 nm ($J_{1.7}$), measured by a PSM. **c**, Particle size distribution versus time, measured by an SMPS. **d**, Gas-phase nitric acid and sulfuric acid versus time, measured by an I⁻ CIMS and a NO₃⁻ CIMS, respectively. Sulfuric acid through SO₂ oxidation started to appear soon after switching on the UV lights at time = 0 min, building up to a steady state of $2.3 \times 10^6 \text{ cm}^{-3}$ after a wall-loss-rate timescale of around 10 min. The subsequent H₂SO₄-NH₃ nucleation led to a relatively slow formation rate of 1.7-nm particles. The particles did not grow above 2.5 nm because of their slow growth rate and corresponding low survival probability against wall loss. Following injection of $2.0 \times 10^9 \text{ cm}^{-3}$ nitric acid into the chamber after 115 min, while leaving the production rate of sulfuric acid and the injection rate of ammonia unchanged, we observed a sharp increase in particle formation rate (panel **b**), together with rapid particle growth of 40 nm h⁻¹ (panel **c**). The overall systematic scale uncertainties of $\pm 30\%$ on particle formation rate, $-33\% / +50\%$ on sulfuric acid concentration and $\pm 25\%$ on nitric acid concentration are not shown.

Source data

At $t = 115$ min, we raised the nitric acid concentration to $2.0 \times 10^9 \text{ cm}^{-3}$, through direct injection instead of photochemical production, so that we could independently control the nitric acid and sulfuric acid concentrations. The particle number increased 30-fold and 1,300-fold for particles larger than 1.7 nm and 2.5 nm, respectively. In addition, these newly formed particles grew much more rapidly (40 nm h⁻¹), reaching 20 nm within 30 min. This experiment shows that nitric acid can substantially enhance particle formation and growth rates for fixed levels of sulfuric acid and ammonia.

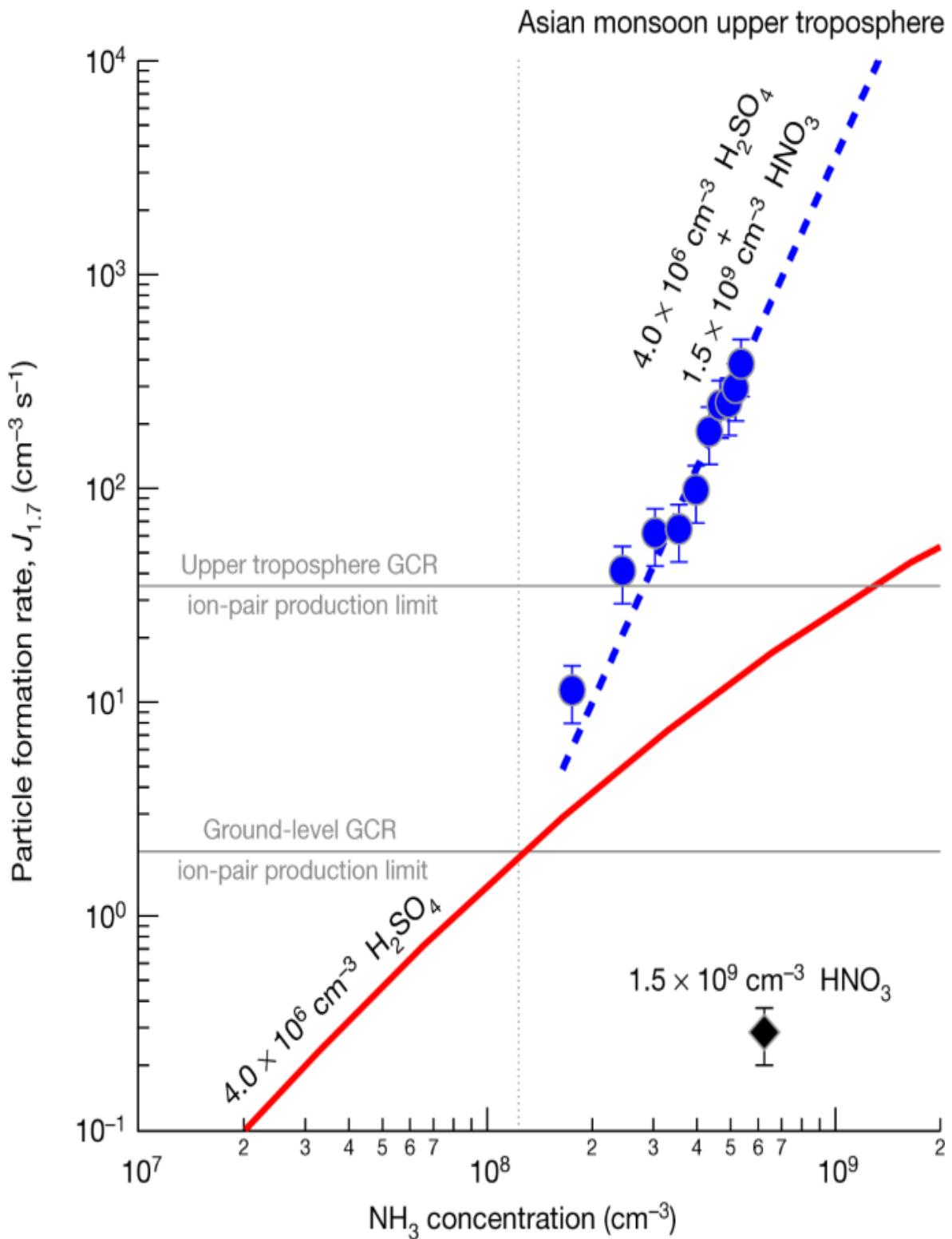
We also conducted model calculations on the basis of known thermodynamics and microphysics ([Methods](#)). Our model results (dashed traces in Fig. [1a](#)) consistently and quantitatively confirm the experimental data: sulfuric acid and ammonia nucleation produces only 1.7-nm particles, whereas addition of nitric acid strongly enhances the formation rates of both 1.7-nm and 2.5-nm particles.

We conducted two further experiments under conditions similar to Fig. [1](#) but holding the concentrations of a different pair of vapours constant while varying the third. For the experiment shown in Extended Data Fig. [1](#), we started by oxidizing NO_2 to produce $1.6 \times 10^9 \text{ cm}^{-3}$ HNO_3 in the presence of about $6.5 \times 10^8 \text{ cm}^{-3}$ NH_3 and then increased H_2SO_4 from 0 to $4.9 \times 10^6 \text{ cm}^{-3}$ by oxidizing progressively more injected SO_2 . For the experiment shown in Extended Data Fig. [2](#), we first established $4.6 \times 10^6 \text{ cm}^{-3}$ H_2SO_4 and $4.0 \times 10^9 \text{ cm}^{-3}$ HNO_3 , and then increased NH_3 from 0 to about $6.5 \times 10^8 \text{ cm}^{-3}$. We consistently observed relatively slow nucleation when only two of the three vapours are present, whereas addition of the third vapour increased nucleation rates by several orders of magnitude.

Figure [2](#) shows particle formation rates measured by CLOUD at 1.7-nm mobility diameter ($J_{1.7}$) versus ammonia concentration, at 223 K. The $J_{1.7}$ data were all measured in the presence of ions from galactic cosmic rays (GCR) and — so — represent the sum of neutral and ion-induced channels. The black diamond shows the measured $J_{1.7}$ of $0.3 \text{ cm}^{-3} \text{ s}^{-1}$ for $\text{HNO}_3\text{--NH}_3$ nucleation with $1.5 \times 10^9 \text{ cm}^{-3}$ nitric acid, about $6.5 \times 10^8 \text{ cm}^{-3}$ ammonia and sulfuric acid below the detection limit of $5 \times 10^4 \text{ cm}^{-3}$ (this is the event shown in Extended Data Fig. [1](#)). At this same ammonia concentration, we measured $J_{1.7} = 6.1 \text{ cm}^{-3} \text{ s}^{-1}$ at $2.3 \times 10^6 \text{ cm}^{-3}$ H_2SO_4 , demonstrating the much faster rate of $\text{H}_2\text{SO}_4\text{--NH}_3$ nucleation (not shown). This measurement is consistent with models on the basis of previous CLOUD studies of $\text{H}_2\text{SO}_4\text{--NH}_3$ nucleation^{[18,19](#)}, as illustrated by the model simulations for $4.0 \times 10^6 \text{ cm}^{-3}$ sulfuric acid (red solid curve). The blue circles show our measurements of $J_{1.7}$ for $\text{HNO}_3\text{--H}_2\text{SO}_4\text{--NH}_3$ nucleation at $4.0 \times 10^6 \text{ cm}^{-3}$ sulfuric acid and $(1.6\text{--}6.5) \times 10^8 \text{ cm}^{-3}$ ammonia, in the presence of 1.5×10^9

cm^{-3} nitric acid (the event shown in Extended Data Fig. 2). The blue dashed curve is a power law fit to the measurements, indicating a strong sensitivity to ammonia concentration $J_{1.7} = k[\{\text{NH}\}]^{3.7}$.

Fig. 2: Particle formation rates at 1.7 nm ($J_{1.7}$) versus ammonia concentration at 223 K and 25% relative humidity.



The chemical systems are $\text{HNO}_3\text{--NH}_3$ (black), $\text{H}_2\text{SO}_4\text{--NH}_3$ (red) and $\text{HNO}_3\text{--H}_2\text{SO}_4\text{--NH}_3$ (blue). The black diamond shows the CLOUD

measurement of $\text{HNO}_3\text{--NH}_3$ nucleation at $1.5 \times 10^9 \text{ cm}^{-3}$ HNO_3 , $6.5 \times 10^8 \text{ cm}^{-3}$ NH_3 and with H_2SO_4 below the detection limit of $5 \times 10^4 \text{ cm}^{-3}$. The red solid curve is $J_{1.7}$ versus ammonia concentration at $4.0 \times 10^6 \text{ cm}^{-3}$ sulfuric acid from a $\text{H}_2\text{SO}_4\text{--NH}_3$ nucleation parameterization on the basis of previous CLOUD measurements^{18,19}. The blue circles show the CLOUD measurements of $\text{HNO}_3\text{--H}_2\text{SO}_4\text{--NH}_3$ nucleation at $4.0 \times 10^6 \text{ cm}^{-3}$ H_2SO_4 , $1.5 \times 10^9 \text{ cm}^{-3}$ HNO_3 and $(1.6\text{--}6.5) \times 10^8 \text{ cm}^{-3}$ NH_3 . The data are fitted by a power law, $J_{1.7} = k[\text{NH}_3]^{3.7}$ (blue dashed curve). The vertical grey dotted line separates ammonia concentrations measured in different regions in the upper troposphere⁵; the region to the right indicates the Asian monsoon conditions. The horizontal grey solid lines show $J_{1.7}$ upper limits for ion-induced nucleation resulting from the GCR ionization rate of around 2 ion pairs $\text{cm}^{-3} \text{ s}^{-1}$ at ground level and 35 ion pairs $\text{cm}^{-3} \text{ s}^{-1}$ in the upper troposphere. Among the three nucleation mechanisms, $\text{H}_2\text{SO}_4\text{--NH}_3$ nucleation dominates in regions with low ammonia (below around $1.0 \times 10^8 \text{ cm}^{-3}$, or 12 pptv), whereas $\text{HNO}_3\text{--H}_2\text{SO}_4\text{--NH}_3$ nucleation dominates at higher ammonia levels characteristic of the Asian monsoon upper troposphere. The bars indicate 30% estimated total error on the particle formation rates. The overall systematic scale uncertainties are $-33\% \text{--} +50\%$ for sulfuric acid and $\pm 25\%$ for nitric acid concentrations.

Source data

The vertical grey dotted line in Fig. 2 separates ammonia concentrations measured in different regions in the upper troposphere⁵; Asian monsoon conditions are to the right of this vertical line. Our results indicate that $\text{H}_2\text{SO}_4\text{--NH}_3$ nucleation is probably responsible for new particle formation in regions with ammonia concentrations below around 10^8 cm^{-3} (12 pptv), but that $\text{HNO}_3\text{--H}_2\text{SO}_4\text{--NH}_3$ nucleation probably dominates at higher ammonia levels in the Asian monsoon upper troposphere. Our nucleation rate measurements confirm that the stronger sulfuric acid is favoured by ammonia in the ammonia-limited regime, so nitric acid will evaporate from the clusters, as it may be displaced by sulfuric acid. However, as ammonia increases from 1.6 to $6.5 \times 10^8 \text{ cm}^{-3}$, we observe sharp increases in $J_{1.7}$ for

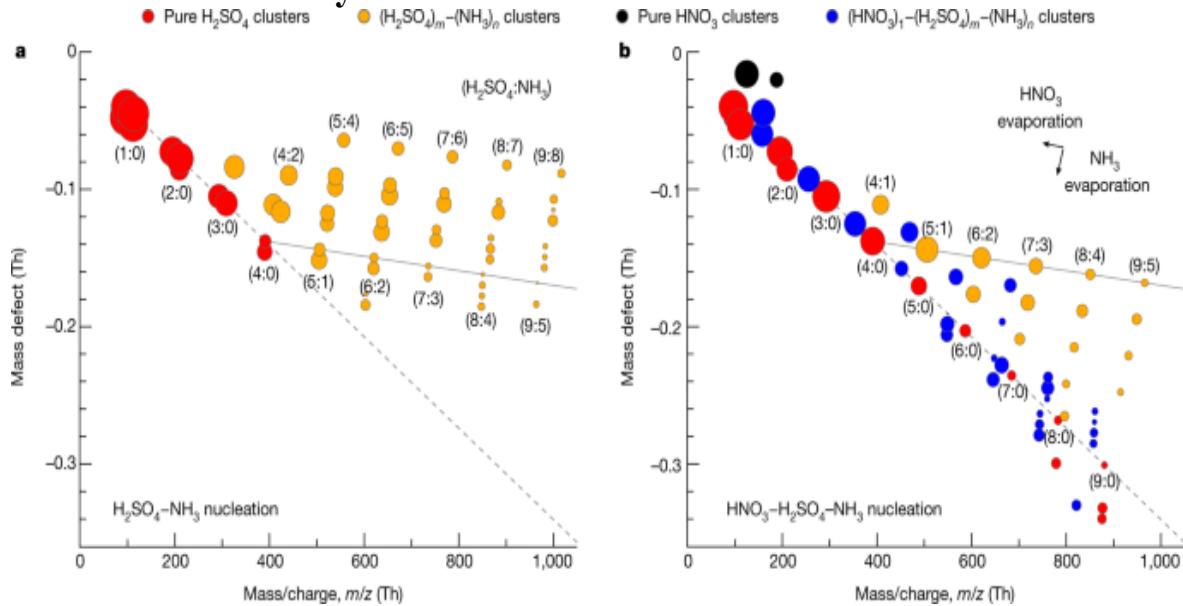
$\text{HNO}_3\text{--H}_2\text{SO}_4\text{--NH}_3$ nucleation from 10 to 400 $\text{cm}^{-3} \text{ s}^{-1}$ and in the ratio of particle formation rates ($\text{HNO}_3\text{--H}_2\text{SO}_4\text{--NH}_3:\text{H}_2\text{SO}_4\text{--NH}_3$) from 4 to 30. Our nucleation model (as in Fig. 1) yields slightly higher $J_{1.7}$ than that observed, as shown in Extended Data Fig. 3, but the formation rate variation with ammonia, nonetheless, shows a similar slope.

CLOUD has previously shown that ions enhance nucleation for all but the strongest acid–base clusters; $\text{HNO}_3\text{--H}_2\text{SO}_4\text{--NH}_3$ is probably not an exception. However, the ion enhancement is limited by the GCR ion-pair production rate. We show with the horizontal grey solid lines in Fig. 2 the upper limits on $J_{1.7}$ for ion-induced nucleation of about $2 \text{ cm}^{-3} \text{ s}^{-1}$ at ground level and $35 \text{ cm}^{-3} \text{ s}^{-1}$ in the upper troposphere. Our experimental nucleation rates for $\text{HNO}_3\text{--H}_2\text{SO}_4\text{--NH}_3$ are mostly above upper tropospheric GCR ion production rates. This is confirmed by similar $J_{1.7}$ measured during a neutral nucleation experiment, in which an electric field was used to rapidly sweep ions from the chamber. Thus, for this nucleation scheme, the neutral channel will often prevail over the ion-induced channel in the Asian monsoon upper troposphere. However, when ammonia is diluted away outside the Asian monsoon anticyclone, ions may enhance the nucleation rate up to the GCR limit near $35 \text{ cm}^{-3} \text{ s}^{-1}$.

In a formal sense, the new-particle-formation mechanism could be one of two types: formation of stable $\text{H}_2\text{SO}_4\text{--NH}_3$ clusters, followed by nano-Köhler-type activation by nitric acid and ammonia¹⁶; or else true synergistic nucleation of nitric acid, sulfuric acid and ammonia⁹. In a practical sense, it makes little difference because coagulation loss is a major sink for all small clusters in the atmosphere²⁰, so appearance of 1.7-nm particles by means of any mechanism constitutes new particle formation. Regardless, we can distinguish between these two possibilities from our measurements of the molecular composition of negatively charged clusters using an atmospheric pressure interface time-of-flight (APi-TOF) mass spectrometer. In Fig. 3, we show cluster mass defect plots during $\text{H}_2\text{SO}_4\text{--NH}_3$ and $\text{HNO}_3\text{--H}_2\text{SO}_4\text{--NH}_3$ nucleation events at 223 K. The marked difference between Fig. 3a, b indicates that nitric acid changes the composition of the nucleating clusters

down to the smallest sizes; thus, the mechanism is almost certainly synergistic $\text{HNO}_3\text{--H}_2\text{SO}_4\text{--NH}_3$ nucleation.

Fig. 3: Molecular composition of negatively charged clusters during $\text{H}_2\text{SO}_4\text{--NH}_3$ and $\text{HNO}_3\text{--H}_2\text{SO}_4\text{--NH}_3$ nucleation events at 223 K and 25% relative humidity.



Mass defect (difference from integer mass) versus mass/charge (m/z) of negatively charged clusters measured with an API-TOF mass spectrometer for $1.7 \times 10^6 \text{ cm}^{-3}$ sulfuric acid and $6.5 \times 10^8 \text{ cm}^{-3}$ ammonia (**a**) and $2.0 \times 10^7 \text{ cm}^{-3}$ sulfuric acid, $3.2 \times 10^9 \text{ cm}^{-3}$ nitric acid and $7.9 \times 10^9 \text{ cm}^{-3}$ ammonia (**b**). The symbol colours indicate the molecular composition as shown. The symbol area is proportional to the logarithm of signal rate (counts per second). The labels ($m:n$) near the symbols indicate the number of sulfuric acid ($\text{H}_2\text{SO}_4)_m$ and ammonia ($\text{NH}_3)_n$ molecules in the clusters, including both neutral and charged species. The grey dashed lines follow clusters that contain pure H_2SO_4 molecules with an HSO_4^- ion (or SO_4^- instead of H_2SO_4 and/or SO_4^- instead of HSO_4^- for pure H_2SO_4 clusters falling below this line in **b**). The grey solid lines follow the 1:1 $\text{H}_2\text{SO}_4\text{--NH}_3$ addition starting at $(\text{H}_2\text{SO}_4)_4\text{--}(\text{NH}_3)_0$. Nearly all clusters in panel **a** lie above this line, whereas nearly all clusters in panel **b** fall below it. Most clusters containing HNO_3 lack NH_3 by the time they are measured (they fall near the ($m:0$) grey dashed line), but the marked difference between **a** and **b** indicates

that the nucleating clusters had distinctly different compositions, probably including relatively weakly bound $\text{HNO}_3\text{--NH}_3$ pairs in **b**. It is probable that nucleating clusters in the CLOUD chamber at 223 K contain $\text{HNO}_3\text{--H}_2\text{SO}_4\text{--NH}_3$ with a roughly 1:1 acid–base ratio. However, during the transmission from the chamber to the warm API-TOF mass spectrometer at 293 K, the clusters lose HNO_3 and NH_3 , leaving a less volatile core of H_2SO_4 with depleted NH_3 . The evaporation of a single NH_3 or HNO_3 molecule from a cluster displaces it on the mass defect plot by a vector distance indicated by the black arrows in **b**.

Source data

In Fig. 3a, the predominant ions are one of several deprotonated sulfuric acid species, including HSO_4^- , SO_4^- , HSO_5^- , SO_5^- and so on, resulting in a group of points for clusters with similar molecular composition but different mass and mass defect. In the figure, we use the labels $(m:n)$ to indicate the number of sulfuric acid and ammonia molecules in the $(\text{H}_2\text{SO}_4)_m\text{--}(\text{NH}_3)_n$ clusters, including both neutral and charged species. The mass defect plot closely resembles those previously measured for $\text{H}_2\text{SO}_4\text{--NH}_3$ nucleation²¹. Negative-ion-induced nucleation proceeds with the known acid–base stabilization mechanism, in which sulfuric acid dimers form as a first step (with HSO_4^- serving as a conjugate base for the first H_2SO_4) and then clusters subsequently grow by 1:1 $\text{H}_2\text{SO}_4\text{--NH}_3$ addition (that is, as ammonium bisulfate)⁹. We use a grey line to illustrate the 1:1 addition path, beginning at $(\text{H}_2\text{SO}_4)_4\text{--}(\text{NH}_3)_0$. Clusters larger than the sulfuric acid tetramers mostly contain several ammonia molecules and, so nearly all clusters in Fig. 3a lie above the grey line.

Figure 3b shows a pronounced change in the cluster API-TOF signal during $\text{HNO}_3\text{--H}_2\text{SO}_4\text{--NH}_3$ nucleation. In addition to pure $(\text{H}_2\text{SO}_4)_m\text{--}(\text{NH}_3)_n$ clusters, we observe clusters with one extra HNO_3 molecule (or NO_3^- ion), that is, $(\text{HNO}_3)_1\text{--}(\text{H}_2\text{SO}_4)_m\text{--}(\text{NH}_3)_n$, and the pure nitric acid monomer and dimer. In sharp contrast with Fig. 3a, all these clusters are deficient in NH_3 , falling below the same grey line as in Fig. 3a. The most deficient contain up to nine bare acids, that is, $(\text{H}_2\text{SO}_4)_9$ or $(\text{H}_2\text{SO}_4)_8\text{--}(\text{HNO}_3)_1$. Figure 3b almost

certainly does not represent the true cluster composition in the chamber because binary nucleation of H_2SO_4 does not proceed under these exact conditions of H_2SO_4 , NH_3 , temperature and relative humidity (as demonstrated by Fig. 3a). We can interpret Fig. 3b as follows. It is probable that clusters in the CLOUD chamber (223 K) contain $\text{HNO}_3-\text{H}_2\text{SO}_4-\text{NH}_3$ with a roughly 1:1 acid–base ratio, representing partial neutralization. However, during the transmission from the cold chamber to the warm API-TOF mass spectrometer (about 293 K), the clusters lose relatively weakly bound HNO_3 and NH_3 molecules but not the lower-volatility H_2SO_4 molecules. Regardless of the interpretation, however, the notable difference between Fig. 3a,b indicates that the sampled clusters had very different compositions and that nitric acid participated in the formation of clusters as small as a few molecules.

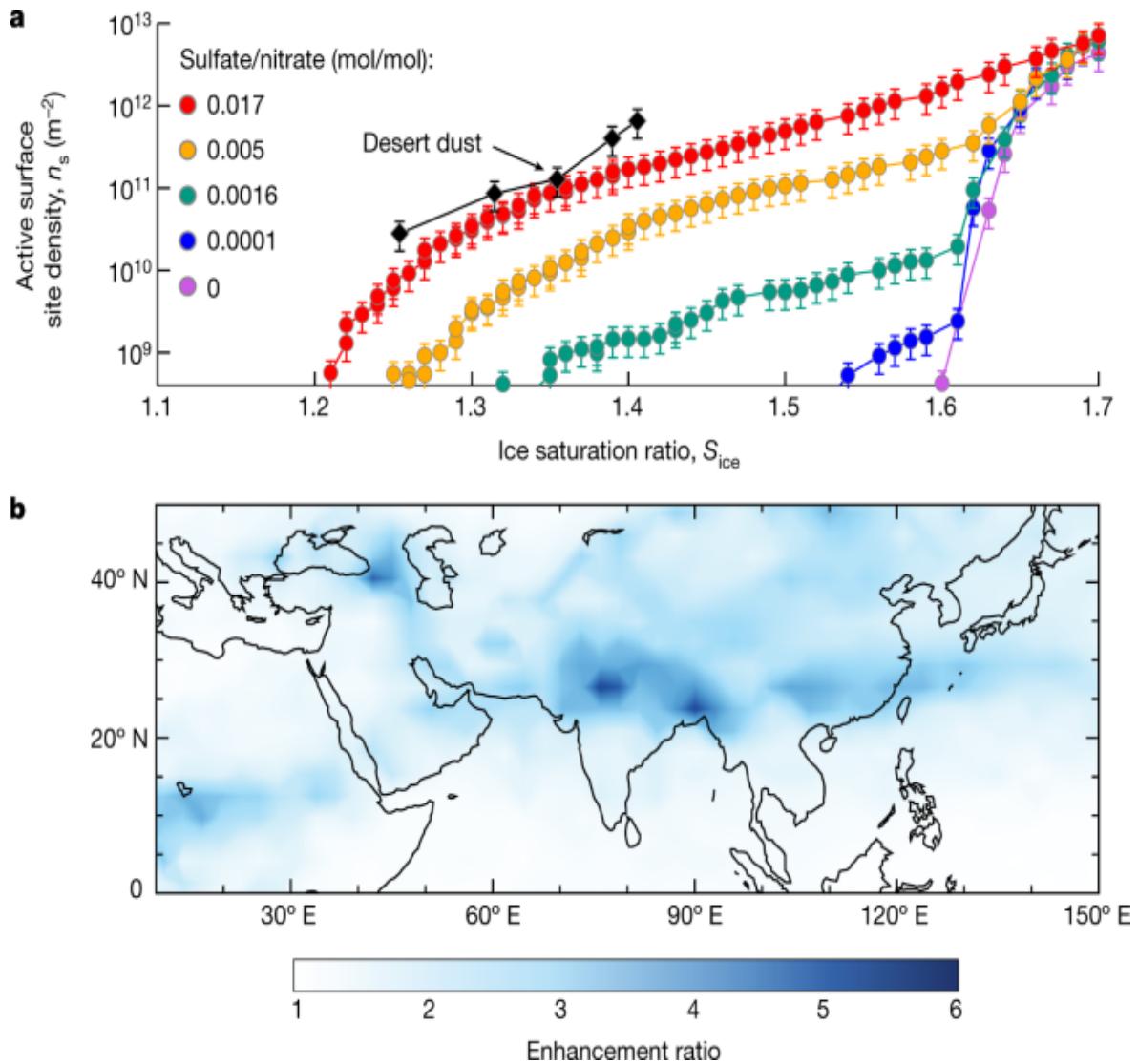
Ice nucleation measurements

Nitric acid and ammonia not only enhance the formation rate of new particles but also drive their rapid growth to sizes at which they may act as CCN or ice nucleating particles (INP), above around 50 nm. To assess their effect on cirrus clouds, we measured the ice nucleation ability of particles formed from $\text{HNO}_3-\text{H}_2\text{SO}_4-\text{NH}_3$ nucleation in the CLOUD chamber.

Simulating ‘hotspot’ conditions, we first formed pure ammonium nitrate particles by means of HNO_3-NH_3 nucleation and then increased the H_2SO_4 fraction in the particles by oxidizing progressively more SO_2 . We measured their ice nucleation ability using the online continuous flow diffusion instrument, mINKA ([Methods](#) and Extended Data Fig. 4). As shown in Fig. 4a, pure ammonium nitrate particles (purple data points) nucleate ice only at high ice saturation ratios (S_{ice}), characteristic of homogeneous nucleation (shown by a steep increase of ice activation above $S_{\text{ice}} = 1.60$ at 215 K). This indicates that pure ammonium nitrate particles, formed by means of HNO_3-NH_3 nucleation, are probably in a liquid state initially, albeit at a relative humidity below the deliquescence point²². However, addition of sulfate, with a particulate sulfate-to-nitrate molar ratio as small as 10^{-4} , triggers crystallization of ammonium nitrate. For these particles, we observed a small heterogeneous ice nucleation mode at S_{ice} of 1.54 (blue data points), with

other conditions and the particle size distribution held almost constant. Moreover, as the sulfate molar fraction progressively rises to just 0.017 (still almost pure but now solid ammonium nitrate), an active surface site density (n_s) of 10^{10} m^{-2} is reached at S_{ice} as low as 1.26. This is consistent with previous findings, in which particles were generated through nebulization, with a much larger particle diameter and a much higher sulfate-to-nitrate ratio²³. Our measurements show that $\text{HNO}_3\text{--H}_2\text{SO}_4\text{--NH}_3$ nucleation followed by rapid growth from nitric acid and ammonia condensation — which results in low sulfate-to-nitrate ratio — could provide an important source of INP that are comparable with typical desert dust particles at nucleating ice²⁴.

Fig. 4: Ice nucleation properties and modelled regional contribution of upper tropospheric particles formed from $\text{HNO}_3\text{--H}_2\text{SO}_4\text{--NH}_3$ nucleation.



a, Active surface site density versus ice saturation ratio, measured by the mINKA instrument at CLOUD, at 233 K and 25% relative humidity. Pure ammonium nitrate particles (purple points) show homogeneous freezing. However, addition of only small amounts of sulfate creates highly ice-nucleation-active particles. At around 1.7% sulfate fraction (red points), the ice nucleating efficiency is comparable with desert dust particles²⁴. **b**, Simulation of particle formation in a global model (EMAC) with efficient vertical transport of ammonia into the upper troposphere during the Asian monsoon. Including multi-acid $\text{HNO}_3\text{--H}_2\text{SO}_4\text{--NH}_3$ nucleation (on the basis of the blue dashed curve in Fig. 2) enhances particle number concentrations (nucleation mode) over the Asian monsoon region by a factor of 3–5

compared with the same model with only $\text{H}_2\text{SO}_4-\text{NH}_3$ nucleation (from Dunne et al.⁸, similar to the red solid curve in Fig. 2).

[Source data](#)

Atmospheric implications

Our findings suggest that $\text{HNO}_3-\text{H}_2\text{SO}_4-\text{NH}_3$ nucleation may dominate new particle formation in the Asian monsoon region of the upper troposphere, with a ‘flame’ of new particles in the outflow of convective clouds, in which up to 10^{10} cm^{-3} ammonia⁶ mixes with low (background) levels of sulfuric acid and nitric acid. Without this mechanism, particle formation through the traditional ternary $\text{H}_2\text{SO}_4-\text{NH}_3$ nucleation would be much slower and most probably rate-limited by the scarce sulfuric acid. Furthermore, by co-condensing with nitric acid, the convected ammonia also drives the growth of the newly formed particles. Given typical acid-excess conditions in the upper troposphere, condensational growth is governed by the availability of ammonia. Consequently, particles will steadily (and rapidly) grow until ammonia is depleted after several *e*-folding times set by the particle condensation sink. On the basis of condensation sinks generally observed in the tropical upper troposphere⁴, this timescale will be several hours. Within this time interval, given the observed ammonia levels, newly formed particles will be able to grow to CCN sizes and even small admixtures of sulfuric acid will render these particles efficient INP.

Our laboratory measurements provide a mechanism that can account for recent observations of abundant ammonium nitrate particles in the Asian monsoon upper troposphere⁶. To evaluate its importance on a global scale, we first parameterized our experimentally measured $J_{1.7}$ for $\text{HNO}_3-\text{H}_2\text{SO}_4-\text{NH}_3$ nucleation as a function of sulfuric acid, nitric acid and ammonia concentrations ([Methods](#)). The parameterization is obtained using a power-law dependency for each vapour (Extended Data Fig. 5), given that the critical cluster composition is associated with the exponents according to the first nucleation theorem²⁵. Then we implemented this parameterization in a global aerosol model (EMAC, see [Methods](#) for modelling details). The EMAC model predicts that $\text{HNO}_3-\text{H}_2\text{SO}_4-\text{NH}_3$ nucleation at 250 hPa (11

km, approximately 223 K) produces an annual average exceeding 1,000 cm⁻³ new particles over an extensive area (Extended Data Fig. 6). This corresponds to an increase in particle number concentration (Fig. 4b) up to a factor of five higher than in a control simulation with only ternary H₂SO₄–NH₃ nucleation⁸. The strongest increase occurs mostly over Asia, in which ammonia is ample because of deep convection from ground sources.

However, another global model (TOMCAT, see [Methods](#)) shows much lower ammonia mixing ratios in the upper troposphere than EMAC (<1 pptv compared with <100 pptv, respectively), although with a broadly similar spatial distribution (Extended Data Fig. 7a,b). This large variability of upper tropospheric ammonia is also indicated by recent field measurements on local^{6,26} and global^{5,27} scales. In view of its importance for both H₂SO₄–NH₃ and HNO₃–H₂SO₄–NH₃ nucleation, there is an urgent need to improve upper tropospheric measurements of ammonia, as well as improve knowledge of its sources, transport and sinks.

We thus turned to a cloud-resolving model to estimate the ammonia vapour fraction remaining after deep convection (see [Methods](#)). We show in Extended Data Fig. 8 that around 10% of the boundary layer ammonia can be transported into the upper troposphere and released as vapour by a base-case convective cloud. The sensitivity tests further illustrate that the key factor governing the fraction of ammonia remaining in the cloud outflow is the retention of ammonia molecules by ice particles (Extended Data Fig. 8e), whereas cloud water pH (Extended Data Fig. 8c) and cloud water content (Extended Data Fig. 8d) only play minor roles once glaciation occurs. Given that more than 10 ppbv of ammonia is often observed in the Asian boundary layer²⁸, it is plausible that the observed 1.4 ppbv (10¹⁰ cm⁻³) ammonia in the upper troposphere⁶ is indeed efficiently transported by the convective systems.

Although the ammonium–nitrate–sulfate particles are formed locally, they can travel from Asia to North America in just three days by means of the subtropical jet stream, as the typical residence time of Aitken mode particles ranges from one week to one month in the upper troposphere²⁹. As a result, these particles can persist as an intercontinental band, covering more than half of the mid-latitude surface area of the Northern Hemisphere (Extended

Data Fig. 6). In summary, synergistic nucleation of nitric acid, sulfuric acid and ammonia could provide an important source of new CCN and ice nuclei in the upper troposphere, especially over the Asian monsoon region, and is closely linked with anthropogenic ammonia emissions²⁷.

Methods

The CLOUD facility

We conducted our measurements at the CERN CLOUD facility, a 26.1-m³, electropolished, stainless-steel CLOUD chamber that allows new-particle-formation experiments under the full range of tropospheric conditions with scrupulous cleanliness and minimal contamination^{9,30}. The CLOUD chamber is mounted in a thermal housing, capable of keeping the temperature constant in the range 208 K and 373 K with a precision of ± 0.1 K (ref. ³¹). Photochemical processes are initiated by homogeneous illumination with a built-in UV fibre-optic system, including four 200-W Hamamatsu Hg-Xe lamps at wavelengths between 250 and 450 nm and a 4-W KrF excimer UV laser at 248 nm with adjustable power. New particle formation under different ionization levels is simulated with and without the electric fields (± 30 kV), which can artificially scavenge or preserve small ions produced from ground-level GCR. Uniform spatial mixing is achieved with magnetically coupled stainless-steel fans mounted at the top and bottom of the chamber. The characteristic gas mixing time in the chamber during experiments is a few minutes. The loss rate of condensable vapours and particles onto the chamber walls is comparable with the ambient condensation sink. To avoid contamination, the chamber is periodically cleaned by rinsing the walls with ultra-pure water and heating to 373 K for at least 24 h, ensuring extremely low contaminant levels of sulfuric acid $< 5 \times 10^4$ cm⁻³ and total organics < 50 pptv (refs. ^{32,33}). The CLOUD gas system is also built to the highest technical standards of cleanliness and performance. The dry air supply for the chamber is provided by boil-off oxygen (Messer, 99.999%) and boil-off nitrogen (Messer, 99.999%) mixed at the atmospheric ratio of 79:21. Highly pure water vapour, ozone and other trace gases such as nitric acid and ammonia can be precisely added at the pptv level from ultra-pure sources.

Instrumentation

Gas-phase sulfuric acid was measured using a nitrate chemical ionization API-TOF (nitrate-CI-API-TOF) mass spectrometer^{34,35} and an iodide chemical ionization time-of-flight mass spectrometer equipped with a Filter Inlet for Gases and Aerosols (I-FIGAERO-CIMS)^{36,37}. The nitrate-CI-API-TOF mass spectrometer is equipped with an electrostatic filter in front of the inlet to remove ions and charged clusters formed in the chamber. A corona charger is used to ionize the reagent nitric acid vapour in a nitrogen flow³⁸. Nitrate ions are then guided in an atmospheric pressure drift tube by an electric field to react with the analyte molecules in the sample flow. Sulfuric acid is quantified for the nitrate-CI-API-TOF with a detection limit of about $5 \times 10^4 \text{ cm}^{-3}$, following the same calibration and loss correction procedures described previously^{9,32,39}. FIGAERO is a manifold inlet for a CIMS with two operating modes. In the sampling mode, a coaxial core sampling is used to minimize the vapour wall loss in the sampling line. The total flow is maintained at 18.0 slpm and the core flow at 4.5 slpm; the CIMS samples at the centre of the core flow with a flow rate of 1.6 slpm. Analyte molecules are introduced into a 150-mbar ion-molecule reactor, chemically ionized by iodide ions that are formed in a Po-210 radioactive source and extracted into the mass spectrometer. The sulfuric acid calibration coefficient for the I-FIGAERO-CIMS is derived using the absolute sulfuric acid concentrations measured with the pre-calibrated nitrate-CI-API-TOF.

Gas-phase nitric acid was also measured using the I-FIGAERO-CIMS. Nitric acid concentration was quantified by measuring HNO_3/N_2 mixtures with known nitric acid concentrations, following similar procedures described previously¹⁶. The HNO_3/N_2 mixture was sourced from flowing 2 slpm ultra-pure nitrogen through a portable nitric acid permeation tube, at constant 40 °C. The permeation rate of nitric acid was determined by passing the outflow of the permeation tube through an impinger containing deionized water and analysing the resulting nitric acid solution through spectrophotometry.

Gas-phase ammonia was either measured or calculated. We measured ammonia using a proton transfer reaction time-of-flight mass spectrometer (PTR3-TOF-MS, or PTR3 for short)⁴⁰. As a carrier gas for the primary ions,

we used argon (ultra-high purity 5.0) to ensure that ammonium ions could not be artificially formed in the region of the corona discharge. Although the theoretical detection limit from peak height and width would be even smaller, the lowest concentration we were able to measure during the first fully ammonia-free runs of the beginning of the campaign was 10^9 cm^{-3} . An explanation for this is that, when concentrations of ammonia are low, effects of wall interaction of the highly soluble ammonia become important and the decay of ammonia in the inlet line becomes very slow. To reduce inlet wall contacts, we used a core-sampling technique directly in front of the instrument to sample only the centre 2 slpm of the 10 slpm inlet flow, but owing to frequent necessary on-site calibrations of volatile organic compounds, a Teflon ball valve was placed within the sample line that probably influenced measurements during times of low ammonia concentrations. At concentrations above about $2 \times 10^9 \text{ cm}^{-3}$ ammonia, however, the response of the instrument was very fast, so that, for example, changes in the chamber ammonia flow rate were easily detectable. Off-site calibrations showed a humidity-independent calibration factor of 0.0017 ncps/ppb. Calibrated data from the PTR3 agree very well with the Picarro above 10^{10} cm^{-3} (detection limit of the Picarro). The PTR3 also provides information about the overall cleanliness of the volatile organic compounds in the chamber. The technique was extensively described previously⁴⁰.

For ammonia concentrations below 10^9 cm^{-3} , we calculated concentration using the calibrated ammonia injection flow and an estimated first-order wall-loss rate. The wall-loss rate (k_{wall}) for ammonia inside the CLOUD chamber is confirmed to be faster than for sulfuric acid⁴¹, and can be determined from the following expression⁴²:

$$\begin{aligned} \$\$ \{ k \}_{\{\{\text{wall}\}\}} = & \frac{A}{V} \frac{2}{\pi} \sqrt{k_e / (D_i)} = \\ & \{ C \}_{\{\{\text{wall}\}\}} \sqrt{D_i} \end{aligned} \quad (1)$$

in which A/V is the surface-to-volume ratio of the chamber, k_e is the eddy diffusion constant (determined by the turbulent mixing intensity, not the transport properties of the gases) and D_i is the diffusion coefficient for each gas. C_{wall} is thus referred to as an empirical parameter of experiment

conditions in the chamber. Here we first determine the k_{wall} for sulfuric acid and nitric acid to be 1.7×10^{-3} and $1.9 \times 10^{-3} \text{ s}^{-3}$, respectively, by measuring their passive decay rates and subtracting the loss rate of chamber dilution for both ($1.2 \times 10^{-3} \text{ s}^{-1}$), as well as the loss rate of dimer formation for sulfuric acid (around $1.6 \times 10^{-3} \text{ s}^{-1}$ for $5 \times 10^6 \text{ cm}^{-3} \text{ H}_2\text{SO}_4$). The k_{wall} for sulfuric acid agrees with our measurements from previous campaigns⁴³. We then derive the C_{wall} for sulfuric acid and nitric acid both to be $2.0 \times 10^{-4} \text{ torr}^{-0.5} \text{ cm}^{-1} \text{ s}^{-0.5}$, with $\langle D \rangle_{\langle \langle \langle \langle \text{H} \rangle \rangle \rangle \langle 2 \rangle \langle \langle \text{SO} \rangle \rangle \langle 4 \rangle} \rangle$ of 74 torr $\text{cm}^2 \text{ s}^{-1}$ and $\langle D \rangle_{\langle \langle \langle \langle \text{HNO} \rangle \rangle \rangle \langle 3 \rangle} \rangle$ of 87 torr $\text{cm}^2 \text{ s}^{-1}$ (ref. 44). Finally, we calculate the k_{wall} for ammonia to be $2.7 \times 10^{-3} \text{ s}^{-1}$, with $\langle D \rangle_{\langle \langle \langle \langle \text{NH} \rangle \rangle \rangle \langle 3 \rangle} \rangle$ of 176 torr $\text{cm}^2 \text{ s}^{-1}$ (ref. 44). Ammonia desorption from the chamber surface is a strong function of the temperature and is believed to be negligible at low temperatures³⁰. Even after a long time exposure, ammonia desorption should be less than $1.6 \times 10^6 \text{ cm}^{-3}$, according to previous parameterization of ammonia background contamination in the CLOUD chamber⁴¹.

The composition of negatively charged ions and clusters were determined using an APi-TOF mass spectrometer⁴⁵. The APi-TOF mass spectrometer is connected to the CLOUD chamber by means of a 1-inch (21.7-mm inner diameter) sampling probe, with coaxial core sampling to minimize the wall losses in the sampling line. The total sample flow is maintained at 20 slpm and the core sample flow for the APi-TOF mass spectrometer at 0.8 slpm. Because this instrument only measures charged clusters, the measurements were made during GCR conditions. Owing to a large temperature difference between the cold chamber (223 K) and the warm APi-TOF mass spectrometer (around 293 K), $\text{HNO}_3\text{--H}_2\text{SO}_4\text{--NH}_3$ clusters probably lose relatively weakly bonded HNO_3 and NH_3 molecules. This resembles the chemical ionization process of detecting ammonia with the nitrate-Cl-APi-TOF, in which HNO_3 and NH_3 molecules rapidly evaporate from the resulting ammonia nitrate cluster in the CI-APi-TOF vacuum regions⁴⁶.

Gas monitors were used to measure ozone (O_3 , Thermo Environmental Instruments TEI 49C), sulfur dioxide (SO_2 , Thermo Fisher Scientific Inc. 42i-TLE) and nitric oxide (NO , ECO Physics, CLD 780TR). Nitrogen

dioxide (NO_2) was measured by a cavity attenuated phase shift nitrogen dioxide monitor (CAPS NO_2 , Aerodyne Research Inc.) and a home-made cavity enhanced differential optical absorption spectroscopy (CE-DOAS) instrument. The relative humidity of the chamber was determined by dew point mirrors (EdgeTech).

Particle number concentrations were monitored by condensation particle counters (CPCs), including an Airmodus A11 nano Condensation Nucleus Counter (nCNC), consisting of a particle size magnifier (PSM) and a laminar-flow butanol-based CPC⁴⁷, as well as a butanol TSI 3776 CPC. Particle size distributions between 1.8 nm and 500 nm were measured by a nano-scanning electrical mobility spectrometer (nSEMS), a nano-scanning mobility particle sizer (nano-SMPS) and a long-SMPS. The nSEMS used a new, radial opposed migration ion and aerosol classifier (ROMIAC), which is less sensitive to diffusional resolution degradation than the DMAs⁴⁸, and a soft X-ray charge conditioner. After leaving the classifier, particles were first activated in a fast-mixing diethylene glycol stage⁴⁹ and then counted with a butanol-based CPC. The nSEMS transfer function that was used to invert the data to obtain the particle size distribution was derived using 3D finite element modelling of the flows, electric field and particle trajectories^{50,51}. The two commercial mobility particle size spectrometers, nano-SMPS and long-SMPS, have been fully characterized, calibrated and validated in several previous studies^{52,53,54}.

Particle-phase chemical composition was quantified using a high-resolution time-of-flight aerosol mass spectrometer (HR-ToF-AMS, Aerodyne Research). The working principles of the HR-ToF-AMS have been explained in detail previously^{55,56}. In brief, particles are focused by an aerodynamic lens and flash-vaporized by impact onto a hot surface at 600 °C under a high vacuum. The vapours are then ionized by 70-eV electrons and the ions are detected with a ToF mass spectrometer. Ionization efficiency calibrations were conducted before and after the campaign and the variation is within 30%. The particle collection efficiency was considered constant during the experiments because temperature and relative humidity in the chamber were fixed and the particle composition was dominated by ammonium nitrate.

INP were measured in real time at 215 K, as a function of ice saturation ratio (S_{ice}), by the mobile ice nucleation instrument of the Karlsruhe Institute of Technology (mINKA). mINKA is a continuous flow diffusion chamber with vertical cylindrical geometry⁵⁷, on the basis of the design of INKA^{58,59}. A detailed description of the continuous flow diffusion chamber working principle is presented elsewhere⁵⁷. Here, predefined scans of the water vapour saturation ratios were performed in the diffusion chamber every 30 min. For each scan, S_{ice} steadily increased from 1.2 to 1.8 while the temperature was kept constant. The errors associated to temperature and S_{ice} inside the diffusion chamber were derived from the uncertainty of the thermocouples attached to the instrument walls (± 0.5 K)⁵⁹.

Determination of particle formation rate

The particle formation rate, $J_{1.7}$, is determined at 1.7-nm mobility diameter (1.4-nm physical diameter), here using a PSM. At 1.7 nm, a particle is normally considered to be above its critical size and, therefore, thermodynamically stable. $J_{1.7}$ is calculated using the flux of the total concentration of particles growing past a specific diameter (here at 1.7 nm), as well as correction terms accounting for aerosol losses owing to dilution in the chamber, wall losses and coagulation. Details were described previously⁴⁷.

Nucleation model

The nucleation model is on the basis of the thermodynamic model for H₂SO₄–NH₃ nucleation described in detail previously^{18,19}. It is developed from the general dynamic equations⁶⁰, to calculate the production and losses for each cluster/particle size to determine the formation rates of the acid–base clusters. For HNO₃–H₂SO₄–NH₃ nucleation, we simplify the model simulations by extrapolating nano-Köhler-type activation by nitric acid and ammonia to clusters down to sulfuric acid trimers. Eighty size bins, ranging from one ammonium sulfate cluster to 300 nm, are used to capture the evolution of the size and composition of polydisperse particles.

In brief, we calculate the equimolar condensation flux of nitric acid and ammonia on the basis of the supersaturation of gas-phase nitric acid and ammonia over particle-phase ammonium nitrate^{[39,60](#)}:

$$\text{\$}\$\{\backslash\text{Phi}\}_{\text{i}}^{\text{v}}=\text{k}_{\text{c}}\{\backslash\text{rm}\{\text{c}\}\}\backslash[\{\text{C}\}_{\text{i}}^{\text{v}}-\text{a}_{\text{i}}\}, \\ \{\text{C}\}_{\text{i}}^{\text{0}}]\text{\$}\$ \\ (2)$$

in which $\langle\langle\backslash\text{Phi}\rangle\rangle_{\text{i}}^{\text{v}}$ is the net condensation flux of nitric acid or ammonia, with vapour concentration $\langle\langle\text{C}\rangle\rangle_{\text{i}}^{\text{v}}$ and saturation concentration $\langle\langle\text{C}\rangle\rangle_{\text{i}}^{\text{0}}$. The term a_i is the activity of species i at the condensed-phase surface of the particle and k_c is the condensation sink for vapours resulting from interaction with particles. The saturation concentrations of nitric acid and ammonia are estimated on the basis of the dissociation constant K_p (ref. [60](#)). When the vapours are unsaturated, particle-phase ammonium nitrate will evaporate to nitric acid and ammonia to reach the equilibrium.

We also include the Kelvin term ($K_{i,p}$) in the simulation to account for the increased activity $\langle\langle\text{a}\rangle\rangle_{\text{i}}=\text{a}'\langle\langle\text{K}\rangle\rangle_{\text{i},\text{p}}$) of a small curved cluster/particle:

$$\text{\$}\$\{\text{K}\}_{\text{i},\text{p}}=\text{10}^{\langle\langle\text{d}\rangle\rangle_{\text{K}}\text{10}}/\langle\langle\text{d}\rangle\rangle_{\text{p}}\text{\$}\$ \\ (3)$$

in which $K_{i,p}$ scales with a ‘Kelvin diameter’ (d_{K10}) for decadal change and d_p is the diameter of the small cluster/particle. The Kelvin diameter for ammonium nitrate is estimated to be 5.3 nm by fitting the data from previous CLOUD experiments according to:

$$\text{\$}\$S=\text{10}^{\langle\langle\text{d}\rangle\rangle_{\text{K10}}/\langle\langle\text{d}\rangle\rangle_{\text{act}}}\text{\$}\$ \\ (4)$$

in which S is the saturation ratio, calculated by means of dividing the product of measured concentrations of nitric acid and ammonia by the dissociation constant K_p and d_{act} is the activation diameter, at which the

thermodynamic energy barrier for condensation is overcome and particles start to grow rapidly.

Determination of ice nucleation ability

During the experiments, aerosol particles were continuously sampled from the CLOUD chamber into the mINKA ice nucleation instrument, using an actively cooled sampling line for a consistent temperature profile. Particles were then subject to well-controlled ice supersaturated conditions; the ones that nucleated ice were selectively detected and counted by an optical particle counter (custom-modified Climet CI-3100, lower detection limit of about 1 μm) located at the outlet of the instrument. Background ice crystals were quantified before each saturation scan (for 2 min) and subtracted from the total ice number concentration of the corresponding measurement. The fraction of INP (f_{ice}) was calculated as the ratio of ice crystals number concentration to the total number of particles larger than 10 nm in diameter. The ice nucleation active surface site density (n_s)⁶¹ was calculated as the ratio of ice number concentration to the total surface area of particles larger than 10 nm in diameter. The overall uncertainty of n_s is estimated to be $\pm 40\%$ (ref. ²⁴). Particle number and surface area concentrations were measured by the SMPS described in the ‘Instrumentation’ section.

In Extended Data Fig. 4, we provide a detailed summary of the measurement data recorded during the ‘hotspot condition’ experiment shown in Fig. 4a, in which we investigated the heterogeneous crystallization and ice nucleation ability of ammonium nitrate/sulfate particles produced directly from new particle formation. We first formed pure ammonium nitrate particles through nucleation of nitric acid and ammonia vapours at 223 K and 15–30% relative humidity (over liquid water). When the evolution of the particle size distribution (Extended Data Fig. 4a) levelled off at a median diameter of around 100 nm, we turned on the UV lights and progressively injected SO₂ at 03:33 to gradually increase sulfuric acid concentration (Extended Data Fig. 4b). Consequently, in Extended Data Fig. 4c, aerosol mass spectrometer measurements show that particle composition was dominated by ammonium nitrate over the course of the experiment, whereas sulfate appeared approximately 1 h after the injection of SO₂. Finally, we show ice nucleation

measurements in Extended Data Fig. 4d. Each vertical trajectory represents a saturation ratio scan in mINKA, colour-coded by the measured ice active fraction (f_{ice}). In each scan, we use a horizontal black dash to indicate an ice onset threshold corresponding to f_{ice} of 10^{-3} . Circles indicate the corresponding scans shown in Fig. 4a.

When the particulate sulfate-to-nitrate molar ratio is smaller than 0.0001, the ice nucleation threshold is detected at an ice saturation ratio (S_{ice}) of about 1.6, consistent with the homogeneous freezing threshold of aqueous solution droplets⁶². This finding shows that, if particles presented as absolutely pure ammonium nitrate (NH_4NO_3), they would exist as supercooled liquid droplets even at very low relative humidity, consistent with previous studies^{22,63}. As the particulate sulfate-to-nitrate molar ratio gradually increases to about 0.017, the ice nucleation onset shifts to a lower S_{ice} of 1.2, caused by heterogeneous ice nucleation on crystalline ammonium nitrate particles²³. Crystalline salts are known to be efficient INP at low temperatures when their deliquescence occurs at higher relative humidity compared with the humidity range of their heterogeneous ice nucleation activity⁶⁴. The fact that the addition of sulfate can promote the crystallization of ammonium nitrate has already been observed in previous studies with particles nebulized in large sizes (around 1 μm) from bulk solutions of ammonium nitrate/sulfate^{6,23,65}. But it is evidenced here for the first time in an in situ particle nucleation and crystallization experiment representative of upper tropospheric conditions.

Particle formation rate parameterization

According to the first nucleation theorem for multicomponent systems²⁵, we parameterize the particle formation rates ($J_{1.7}$) for the $\text{HNO}_3\text{--H}_2\text{SO}_4\text{--NH}_3$ nucleation scheme with the empirical formula:

$$\begin{aligned} \$\$ \{ J \}_{1.7} = & k \cdot [\{ \{ \text{rm} \{ H \} \} \}_{2} \{ \{ \text{rm} \{ SO \} \} \}_{4}]^a \cdot \\ & [\{ \{ \text{rm} \{ HNO \} \} \}_{3}]^b \cdot [\{ \{ \text{rm} \{ NH \} \} \}_{3}]^c \$\$ \end{aligned} \quad (5)$$

in which vapour concentrations are in units of cm^{-3} and k , a , b and c are free parameters. This method has been validated by previous observations that the particle formation rates ($J_{1.7}$) vary as a product of power-law functions of nucleating vapours. For example, $J_{1.7}$ for ternary sulfuric acid, ammonia (and water) nucleation follows a cubic dependency on sulfuric acid⁸ and a linear⁸ or quadratic¹⁹ dependency on ammonia; $J_{1.7}$ for multicomponent nucleation of sulfuric acid, biogenic oxidized organics and ammonia follows a quadratic dependency on sulfuric acid, a linear dependency on both organics⁶⁶ and ammonia¹¹. The prefactor k accounts for effects of external conditions, such as temperature and relative humidity, thus differs in different environments.

To isolate variables, here we fit the power-law exponents for sulfuric acid, nitric acid and ammonia, respectively, to the dataset of experiments in which only the corresponding vapour concentration was varied. The red triangles, blue circles and yellow squares in Extended Data Fig. 5a–c (same experiments in Extended Data Fig. 1, Fig. 1 and Extended Data Fig. 2), respectively, show that $J_{1.7}$ depends on $[\text{H}_2\text{SO}_4]^3$ for sulfuric acid between 2.6×10^5 and $2.9 \times 10^6 \text{ cm}^{-3}$ (or 0.008 and 0.09 pptv), on $[\text{HNO}_3]^2$ for nitric acid between 2.3×10^8 and $1.7 \times 10^9 \text{ cm}^{-3}$ (or 7 and 52 pptv) and on $[\text{NH}_3]^4$ for ammonia between 1.7×10^8 and $4.9 \times 10^8 \text{ cm}^{-3}$ (or 5 and 15 pptv). The third power exponent for sulfuric acid is consistent with previously reported parameterizations for ternary $\text{H}_2\text{SO}_4-\text{NH}_3$ nucleation^{8,19}. The fourth power exponent for ammonia, however, is larger than those in ternary^{8,19} or multicomponent systems¹¹, which emphasizes the critical role of ammonia and suggests further bonding between ammonia and nitric acid molecules in the nucleating clusters. Next, we verify the exponents by refitting the product of $[\text{H}_2\text{SO}_4]^3$, $[\text{HNO}_3]^2$ and $[\text{NH}_3]^4$ to the full dataset. Extended Data Fig. 5d shows good consistency ($R^2 = 0.9$) of the parameterization among the three experiments, with a slope of $3.4 \times 10^{-71} \text{ s}^{-1} \text{ cm}^{24}$ being the prefactor k :

$$\begin{aligned} \$\$ \{ J \}_{1.7} = & 3.4 \times 10^{-71} \{ [\{ \{ \text{rm} \{ H \} \} \}_{2} \\ & \{ \{ \text{rm} \{ SO \} \}_{4} \}_{3} \} \{ [\{ \{ \text{rm} \{ HNO \} \}_{3} \}_{2} \} \\ & \{ [\{ \{ \text{rm} \{ NH \} \}_{3} \}_{4} \}_{4} \} \end{aligned} \quad (6)$$

This parameterization is representative of new particle formation in the Asian monsoon upper troposphere because our experimental conditions of 223 K and 25% relative humidity, as well as concentrations of sulfuric acid^{67,68} and nitric acid^{69,70}, are within the upper tropospheric range, with ammonia^{5,6} typical of Asian monsoon regions. One caveat, however, is that the cosmic radiation was at the ground level in our chamber, as shown with grey dot-dashed horizontal line in Extended Data Fig. 5d. The ion-pair production rate can be up to ten times higher in the ambient upper troposphere⁷¹, potentially leading to further enhancement of $J_{1.7}$ by ion-induced nucleation, although the neutral channel dominates in our experiments.

Estimated temperature dependence of the particle formation rate

We did not cover the full temperature range in the upper troposphere, instead focusing on 223 K. However, to make the parameterization in the previous section more applicable for model simulations while not overstating the role of this mechanism, we provide some constraints on the temperature dependence of $J_{1.7}$ for $\text{HNO}_3\text{--H}_2\text{SO}_4\text{--NH}_3$ nucleation.

Broadly, it is certain that particle formation involving HNO_3 will have a strong temperature dependence, becoming much slower as T increases.

We first present the temperature dependence of $J_{1.7}$ for pure $\text{HNO}_3\text{--NH}_3$ nucleation with the expression:

$$\$ \$ \{ J \} _{1.7} = k(T) f([\{ \{ \text{rm} \{ \text{HNO} \} \} _3 \}, [\{ \{ \text{rm} \{ \text{NH} \} \} _3 \}]) \$ \$ \quad (7)$$

in which $k(T)$ is an empirical temperature-dependent rate constant and has the Arrhenius form

$$\$ \$ k(T) = \{ \{ \text{rm} \{ e \} \} \} ^{ \{ \left(-\frac{1}{T} \frac{E}{R} \right) } , \$ \$ \quad (8)$$

in which T is the absolute temperature (in Kelvin), E is the activation energy and R is the universal gas constant. $f([\text{HNO}_3], [\text{NH}_3])$ is a function of the ammonia and nitric acid concentrations (including the pre-exponential factor and free-fitting parameters). This expression is then fitted to the dataset in Fig. 3c in our previous study¹⁶, in which $J_{1.7}$ were measured with only nitric acid, ammonia and water vapours added to the chamber, and the temperature was progressively decreased from 258 K to 249 K. Because the ammonia and nitric acid concentrations were kept almost constant during the temperature transition, we treat the $f([\text{HNO}_3], [\text{NH}_3])$ term as a constant to reduce the degrees of freedom. This expression with its two free parameters leads to a good agreement with the data, $R_2 = 0.96$. And the fitted $-E/R$ and $f([\text{HNO}_3], [\text{NH}_3])$ are 14,000 K and 3.2×10^{-26} , respectively.

Next, we apply the same $k(T)$ term to the $\text{HNO}_3\text{--H}_2\text{SO}_4\text{--NH}_3$ parameterization (equation (9)), assuming that the multicomponent nucleation follows a similar temperature dependence:

$$\text{\$}\{ J \}_{1.7} = 2.9 \times 10^{-98} \left(\frac{14,000}{T} \right)^{\frac{1}{3}} \left[\left(\frac{[\text{H}]}{[\text{SO}_4]} \right)^2 \left(\frac{[\text{HNO}_3]}{[\text{NH}_3]} \right)^3 \right]^{\frac{1}{4}} \text{\$} \quad (9)$$

Although this temperature-dependent parameterization may not be the final description of this process, it tracks the trend of $J_{1.7}$ well. In the event of $4 \times 10^6 \text{ cm}^{-3} \text{ H}_2\text{SO}_4$, $1.5 \times 10^9 \text{ cm}^{-3} \text{ HNO}_3$ and $5 \times 10^8 \text{ cm}^{-3} \text{ NH}_3$, the multicomponent nucleation is quenched ($J_{1.7} < 0.01 \text{ cm}^{-3} \text{ s}^{-1}$) above 268 K. This is consistent with the observations that nitric acid and ammonia only contribute to the growth of ammonium sulfate particles at 278 K (ref. 16). At 223 K, the parameterized $J_{1.7}$ is $306 \text{ cm}^{-3} \text{ s}^{-1}$, matching our measurement in Fig. 2. And for the temperature in the upper troposphere and lower stratosphere (198 K), the parameterized $J_{1.7}$ is $8 \times 10^5 \text{ cm}^{-3} \text{ s}^{-1}$, which is still much slower than its kinetic limit of about $10^9\text{--}10^{10} \text{ cm}^{-3} \text{ s}^{-1}$.

The EMAC global model

The ECHAM/MESSy Atmospheric Chemistry (EMAC) model is a numerical chemistry and climate simulation system that includes sub-models describing tropospheric and middle atmosphere processes and their interaction with oceans, land and human influences⁷². It uses the second version of the Modular Earth Submodel System (MESSy2) to link multi-institutional computer codes. Atmospheric circulation is calculated by the 5th generation of the European Centre Hamburg general circulation model (ECHAM5 (ref. 73)) and atmospheric chemical kinetics are solved for every model time step. For the present study, we applied EMAC (ECHAM5 version 5.3.02, MESSy version 2.54.0) in the T42L31ECMWF-resolution, for example, with a spherical truncation of T42 (corresponding to a quadratic Gaussian grid of approximately 2.8° by 2.8° in latitude and longitude) with 31 vertical hybrid pressure levels up to 10 hPa. EMAC uses

a modal representation of aerosols dynamics (GMXe) that describes the aerosol size distribution as seven interacting log-normal distributions, of which four modes are soluble and three modes are insoluble. New particles are added directly to the nucleation mode. The applied model setup comprises the sub-model New Aerosol Nucleation (NAN) that includes new parameterizations of aerosol particle formation rates published in recent years⁷⁴. These parameterizations include ion-induced nucleation. The ion-pair production rate, needed to calculate the ion-induced or ion-mediated nucleation, is described using the sub-model IONS, which provides ion-pair production rates⁷⁴.

The TOMCAT global model

The TOMCAT model is a global 3D offline chemical transport model^{75,76}. It is run at approximately 2.8° spatial resolution, such as EMAC on a T42 grid, driven by ECMWF ERA-Interim reanalysis meteorological fields for the year 2008. We also used 31 hybrid sigma-pressure levels from the surface to 10 hPa. The dissolved fraction of gases in cloud water is calculated by means of an equilibrium Henry's law approach and set to zero for temperatures below -20°C . The model includes GLOMAP aerosol microphysics⁷⁷ with nitrate and ammonium from the HyDIS solver⁷⁸ and the representation of new particle formation used by Gordon et al.³. The HyDIS solver adopts a sophisticated approach to the dissolution of nitric acid and ammonia into the aerosol phase that is a hybrid between a dynamic representation of the process, which accounts for the time needed for mass transport, and an equilibrium representation, which does not⁷⁸. The main limitation of the solver is that it assumes all aerosol particles are liquid, which is probably a poor approximation in cold, dry conditions frequently found in the upper troposphere.

The cloud trajectories framework

We conducted a sensitivity study on ammonia transport processes and estimated the fraction remaining of ammonia vapour after convection from the boundary layer to the upper troposphere, using a cloud trajectories framework described in detail in Bardakov et al.^{79,80}. In brief, trajectories

from a convective system simulated with the large-eddy simulation (LES) model MIMICA⁸¹ were extracted and a parcel representing the cloud outflow was selected for further analysis (Extended Data Fig. 8a). The meteorological profiles and clouds microphysics scheme used here were the same as in Bardakov et al.⁸⁰, producing altitude-dependent distributions of water and ice hydrometeors depicted in Extended Data Fig. 8. Partitioning of gas between vapour and aqueous phase along the trajectory was calculated on the basis of Henry's law constant adjusted to a cloud pH, $H^* = H \times 1.7 \times 10^{(9-\text{pH})}$ following the expression for ammonia from Seinfeld and Pandis⁶⁰.

We then investigated the factors governing ammonia transport through the simulated convective system by varying: (1) the pH for the liquid water hydrometeors (Extended Data Fig. 8c); (2) the total amount of water in the system (Extended Data Fig. 8d); (3) the retention of ammonia molecules by the ice hydrometeors (Extended Data Fig. 8e). In our base-case simulation, the pH was assumed to have an altitude-dependent profile, reflecting the higher abundance of acids close to the surface and ranging from 4.5 to 5, in accordance with the representative pH values in the EMAC simulation. The base-case water content was as in Bardakov et al.⁸⁰ and the ice retention coefficient 0.05 in accordance with Ge et al.¹³, with no further uptake on ice.

Atmospheric interpretation

This work focuses on the Asian monsoon region in part because this region is fairly extensive, but also because ammonia concentrations measured in this region are by far the highest in the upper troposphere. Although we frame this synergistic $\text{HNO}_3\text{--H}_2\text{SO}_4\text{--NH}_3$ nucleation in a scenario that suits the Asian monsoon upper troposphere, the physics applies more broadly — the colder the conditions are, the more important this mechanism is likely to be. To explore the importance of this synergistic nucleation to the atmosphere, we combine our experimental results, cloud resolving modelling and global-scale chemical transport modelling. On the basis of these constraints, the rate-limiting elements of new particle formation seem to be convective transport of ammonia and the production rate of particles

in the mixing zone between convective outflow and the background upper free troposphere; however, confirmation of this will require extensive field and modelling studies.

Generally, nitric acid ranges between about 10^8 and 10^9 cm^{-3} (refs. [14,15](#)) and sulfuric acid between about 10^5 and 10^6 cm^{-3} (refs. [82,83](#)) in the tropical upper troposphere. The typical acid-excess conditions leave the principal uncertainty being ammonia levels, which are not yet well constrained. Although satellite-based ammonia measurements have provided a spatial distribution on a global scale, they are limited to cloud-free areas owing to blockage of the ammonia signal by optically thick clouds. However, deep convection followed by cloud glaciation may be a major source of upper tropospheric ammonia. This process may then not be captured by satellites as it occurs near clouds, with short time duration and high spatial heterogeneity. This may also explain why the in situ-measured ammonia concentrations are up to 40 times higher than those from satellite measurements^{[6](#)}.

Ammonia has no known chemical source in the atmosphere but is instead transported by cloud processes from the surface, whereas nitric acid and sulfuric acid vapours are formed primarily by out-of-cloud oxidation. Consequently, it is probable that this synergistic nucleation occurs initially in the outflow of convective clouds, in which the released ammonia mixes with pre-existing (background) nitric acid and sulfuric acid. Subsequently, as ammonia is titrated over several *e*-folding times (governed by the condensation sink in this mixing zone) and the outflow air fully mixes with the background air, nucleation conditions will shift from the ammonia-rich regime to the ammonia-limited regime. These highly dynamic processes are thus the key to constraining the climatic effects of this synergistic nucleation in Asian monsoon and potentially other convective regions. Nevertheless, current ambient measurements confirm the presence of ample ammonia, as well as particles comprised largely of ammonium nitrate^{[4](#)}, and our experiments show that synergistic $\text{HNO}_3\text{--H}_2\text{SO}_4\text{--NH}_3$ nucleation is a viable mechanism for new particle formation in the Asian monsoon upper troposphere. As global ammonia emissions continue to increase owing to agricultural growth and the warmer climate^{[84,85](#)}, the importance of this particle formation mechanism will increase.

Further, as there is almost no in situ composition measurement of clusters or newly formed particles in the upper troposphere, we can only infer the major particle formation pathway from indirect evidence such as composition of precursor vapours or larger particles. Previously established mechanisms include binary and ternary sulfuric acid nucleation, which drive new particle formation over marine or anthropogenically influenced regions^{1,4,86,87}, nucleation by oxygenated organics, which dominates over pristine vegetated areas such as the Amazon basin^{2,10,88}, and nucleation by iodine oxidation products, which may be especially important in marine convection^{89,90}. Over the Asian monsoon regions, however, mixed emissions of both inorganic and organic vapours may well complicate the particle formation mechanism. However, it has been demonstrated that ammonium nitrate can often explain more than half of the particulate volume in the upper troposphere⁶. This means that the $\text{HNO}_3\text{--NH}_3$ concentration is probably higher than the sum of all other condensable vapours (that is, sulfuric acid and oxygenated organics). And given that $\text{HNO}_3\text{--H}_2\text{SO}_4\text{--NH}_3$ nucleation is orders of magnitude faster than binary and ternary sulfuric acid nucleation at observed ammonia levels, we therefore infer that synergistic $\text{HNO}_3\text{--H}_2\text{SO}_4\text{--NH}_3$ nucleation is a major particle formation pathway in the Asian monsoon upper troposphere. It seems unlikely that this inorganic pathway and the organic pathways are antagonistic in growth, and without strong indications otherwise, it seems probable that they are more or less additive for nucleation itself. However, to further investigate interactions between different nucleation schemes, we would rely on further information on the source and identity of organic vapours that are present in the Asian monsoon upper troposphere.

Data availability

The full dataset shown in the figures is publicly available at <https://doi.org/10.5281/zenodo.5949440>. [Source data](#) are provided with this paper.

Code availability

The EMAC (ECHAM/MESSy) model is continuously further developed and applied by a consortium of institutions. The use of MESSy and access to the source code is licensed to all affiliates of institutions that are members of the MESSy Consortium. Institutions can become a member of the MESSy Consortium by signing the MESSy Memorandum of Understanding. More information can be found on the MESSy Consortium website (<https://www.messy-interface.org>). All code modifications presented in this paper will be included in the next version of MESSy. The TOMCAT model (<http://homepages.see.leeds.ac.uk/~lecmc/tomcat.html>) is a UK community model. It is available to UK (or NERC-funded) researchers who normally access the model on common facilities or who are helped to install it on their local machines. As it is a complex research tool, new users will need help to use the model optimally. We do not have the resources to release and support the model in an open way. Any potential user interested in the model should contact Martyn Chipperfield. The model updates described in this paper are included in the standard model library. The cloud trajectories model is publicly available at <https://doi.org/10.5281/zenodo.5949440>. Codes for conducting the analysis presented in this paper can be obtained by contacting the corresponding author, Neil M. Donahue (nmd@andrew.cmu.edu).

References

1. Clarke, A. et al. Nucleation in the equatorial free troposphere: favorable environments during PEM-Tropics. *J. Geophys. Res. Atmos.* **104**, 5735–5744 (1999).
2. Weigel, R. et al. In situ observations of new particle formation in the tropical upper troposphere: the role of clouds and the nucleation mechanism. *Atmos. Chem. Phys.* **11**, 9983–10010 (2011).
3. Gordon, H. et al. Causes and importance of new particle formation in the present-day and pre-industrial atmospheres. *J. Geophys. Res. Atmos.* **122**, 8739–8760 (2017).
4. Williamson, C. J. et al. A large source of cloud condensation nuclei from new particle formation in the tropics. *Nature* **574**, 399–403

(2019).

5. Höpfner, M. et al. First detection of ammonia (NH_3) in the Asian summer monsoon upper troposphere. *Atmos. Chem. Phys.* **16**, 14357–14369 (2016).
6. Höpfner, M. et al. Ammonium nitrate particles formed in upper troposphere from ground ammonia sources during Asian monsoons. *Nat. Geosci.* **12**, 608–612 (2019).
7. Intergovernmental Panel on Climate Change. *Climate Change 2013: The Physical Science Basis* (Cambridge Univ. Press, 2013).
8. Dunne, E. M. et al. Global atmospheric particle formation from CERN CLOUD measurements. *Science* **354**, 1119–1124 (2016).
9. Kirkby, J. et al. Role of sulphuric acid, ammonia and galactic cosmic rays in atmospheric aerosol nucleation. *Nature* **476**, 429–433 (2011).
10. Andreae, M. O. et al. Aerosol characteristics and particle production in the upper troposphere over the Amazon Basin. *Atmos. Chem. Phys.* **18**, 921–961 (2018).
11. Lehtipalo, K. et al. Multicomponent new particle formation from sulfuric acid, ammonia, and biogenic vapors. *Sci. Adv.* **4**, eaau5363 (2018).
12. Zhao, B. et al. High concentration of ultrafine particles in the Amazon free troposphere produced by organic new particle formation. *Proc. Natl Acad. Sci.* **117**, 25344–25351 (2020).
13. Ge, C., Zhu, C., Francisco, J. S., Zeng, X. C. & Wang, J. A molecular perspective for global modeling of upper atmospheric NH_3 from freezing clouds. *Proc. Natl Acad. Sci.* **115**, 6147–6152 (2018).
14. Martin, R. V. et al. Space-based constraints on the production of nitric oxide by lightning. *J. Geophys. Res. Atmos.* **112**, D09309 (2007).

15. Lelieveld, J. et al. The South Asian monsoon—pollution pump and purifier. *Science* **361**, 270–273 (2018).
16. Wang, M. et al. Rapid growth of atmospheric nanoparticles by nitric acid and ammonia condensation. *Nature* **580**, 184–189 (2020).
17. Stolzenburg, D. et al. Enhanced growth rate of atmospheric particles from sulfuric acid. *Atmos. Chem. Phys.* **20**, 7359–7372 (2020).
18. Kürten, A. New particle formation from sulfuric acid and ammonia: nucleation and growth model based on thermodynamics derived from CLOUD measurements for a wide range of conditions. *Atmos. Chem. Phys.* **19**, 5033–5050 (2019).
19. Xiao, M. et al. The driving factors of new particle formation and growth in the polluted boundary layer. *Atmos. Chem. Phys.* **21**, 14275–14291 (2021).
20. Ehrhart, S. & Curtius, J. Influence of aerosol lifetime on the interpretation of nucleation experiments with respect to the first nucleation theorem. *Atmos. Chem. Phys.* **13**, 11465–11471 (2013).
21. Schobesberger, S. et al. Molecular understanding of atmospheric particle formation from sulfuric acid and large oxidized organic molecules. *Proc. Natl Acad. Sci.* **110**, 17223–17228 (2013).
22. Martin, S. T. Phase transitions of aqueous atmospheric particles. *Chem. Rev.* **100**, 3403–3454 (2000).
23. Wagner, R. et al. Solid ammonium nitrate aerosols as efficient ice nucleating particles at cirrus temperatures. *J. Geophys. Res. Atmos.* **125**, e2019JD032248 (2020).
24. Ullrich, R. et al. A new ice nucleation active site parameterization for desert dust and soot. *J. Atmos. Sci.* **74**, 699–717 (2017).
25. Oxtoby, D. W. & Kashchiev, D. A general relation between the nucleation work and the size of the nucleus in multicomponent

- nucleation. *J. Chem. Phys.* **100**, 7665–7671 (1994).
26. Kille, N. et al. The CU mobile solar occultation flux instrument: structure functions and emission rates of NH₃, NO₂ and C₂H₆. *Atmos. Meas. Tech.* **10**, 373–392 (2017).
 27. Nault, B. A. et al. Chemical transport models often underestimate inorganic aerosol acidity in remote regions of the atmosphere. *Commun. Earth Environ.* **2**, 93 (2021).
 28. Warner, J. X., Wei, Z., Strow, L. L., Dickerson, R. R. & Nowak, J. B. The global tropospheric ammonia distribution as seen in the 13-year AIRS measurement record. *Atmos. Chem. Phys.* **16**, 5467–5479 (2016).
 29. Williams, J., Reus, M. D., Krejci, R., Fischer, H. & Ström, J. Application of the variability-size relationship to atmospheric aerosol studies: estimating aerosol lifetimes and ages. *Atmos. Chem. Phys.* **2**, 133–145 (2002).
 30. Duplissy, J. et al. Effect of ions on sulfuric acid-water binary particle formation: 2. Experimental data and comparison with QC-normalized classical nucleation theory. *J. Geophys. Res. Atmos.* **212**, 1752–1775 (2016).
 31. Dias, A. et al. Temperature uniformity in the CERN CLOUD chamber. *Atmos. Meas. Tech.* **10**, 5075–5088 (2017).
 32. Kirkby, J. et al. Ion-induced nucleation of pure biogenic particles. *Nature* **530**, 521–526 (2016).
 33. Schnitzhofer, R. et al. Characterisation of organic contaminants in the CLOUD chamber at CERN. *Atmos. Meas. Tech.* **7**, 2159–2168 (2014).
 34. Jokinen, T. et al. Atmospheric sulphuric acid and neutral cluster measurements using CI-APi-TOF. *Atmos. Chem. Phys.* **12**, 4117–4125 (2012).

35. Kürten, A. et al. Neutral molecular cluster formation of sulfuric acid–dimethylamine observed in real time under atmospheric conditions. *Proc. Natl Acad. Sci.* **111**, 15019–15024 (2014).
36. Lopez-Hilfiker, F. D. et al. A novel method for online analysis of gas and particle composition: description and evaluation of a Filter Inlet for Gases and AEROsols (FIGAERO). *Atmos. Meas. Tech.* **7**, 983–1001 (2014).
37. Wang, M. et al. Reactions of atmospheric particulate stabilized Criegee intermediates lead to high-molecular-weight aerosol components. *Environ. Sci. Technol.* **50**, 5702–5710 (2016).
38. Kürten, A., Rondo, L., Ehrhart, S. & Curtius, J. Performance of a corona ion source for measurement of sulfuric acid by chemical ionization mass spectrometry. *Atmos. Meas. Tech.* **4**, 437–443 (2011).
39. Tröstl, J. et al. The role of low-volatility organic compounds in initial particle growth in the atmosphere. *Nature* **530**, 527–531 (2016).
40. Breitenlechner, M. et al. PTR3: an instrument for studying the lifecycle of reactive organic carbon in the atmosphere. *Anal. Chem.* **89**, 5824–5831 (2017).
41. Kürten, A. et al. Experimental particle formation rates spanning tropospheric sulfuric acid and ammonia abundances, ion production rates, and temperatures. *J. Geophys. Res. Atmos.* **121**, 12–377 (2016).
42. McMurry, P. H. & Grosjean, D. Gas and aerosol wall losses in Teflon film smog chambers. *Environ. Sci. Technol.* **19**, 1176–1181 (1985).
43. Simon, M. et al. Molecular insight into HOM formation and biogenic new-particle formation over a wide range of tropospheric temperatures. *Atmos. Chem. Phys.* **20**, 9183–9207 (2020).
44. Tang, M., Cox, R. & Kalberer, M. Compilation and evaluation of gas phase diffusion coefficients of reactive trace gases in the atmosphere:

- volume 1. inorganic compounds. *Atmos. Chem. Phys.* **14**, 9233–9247 (2014).
45. Junninen, H. et al. A high-resolution mass spectrometer to measure atmospheric ion composition. *Atmos. Meas. Tech.* **3**, 1039–1053 (2010).
 46. Kürten, A. et al. Observation of new particle formation and measurement of sulfuric acid, ammonia, amines and highly oxidized organic molecules at a rural site in central Germany. *Atmos. Chem. Phys.* **16**, 12793–12813 (2016).
 47. Dada, L. et al. Formation and growth of sub-3-nm aerosol particles in experimental chambers. *Nat. Protoc.* **15**, 1013–1040 (2020).
 48. Mui, W., Mai, H., Downard, A. J., Seinfeld, J. H. & Flagan, R. C. Design, simulation, and characterization of a radial opposed migration ion and aerosol classifier (ROMIAC). *Aerosol Sci. Technol.* **51**, 801–823 (2017).
 49. Wimmer, D. et al. Performance of diethylene glycol-based particle counters in the sub-3 nm size range. *Atmos. Meas. Tech.* **6**, 1793–1804 (2013).
 50. Mai, H. & Flagan, R. C. Scanning DMA data analysis I. Classification transfer function. *Aerosol Sci. Technol.* **52**, 1382–1399 (2018).
 51. Mai, H., Kong, W., Seinfeld, J. H. & Flagan, R. C. Scanning DMA data analysis II. Integrated DMA-CPC instrument response and data inversion. *Aerosol Sci. Technol.* **52**, 1400–1414 (2018).
 52. Jurányi, Z. et al. A 17 month climatology of the cloud condensation nuclei number concentration at the high alpine site Jungfraujoch. *J. Geophys. Res. Atmos.* **116**, D10204 (2011).
 53. Tröstl, J. et al. Fast and precise measurement in the sub-20 nm size range using a Scanning Mobility Particle Sizer. *J. Aerosol Sci.* **87**, 75–87 (2015).

54. Wiedensohler, A. et al. Mobility particle size spectrometers: harmonization of technical standards and data structure to facilitate high quality long-term observations of atmospheric particle number size distributions. *Atmos. Meas. Tech.* **5**, 657–685 (2012).
55. Jayne, J. T. et al. Development of an aerosol mass spectrometer for size and composition analysis of submicron particles. *Aerosol Sci. Technol.* **33**, 49–70 (2000).
56. DeCarlo, P. F. et al. Field-deployable, high-resolution, time-of-flight aerosol mass spectrometer. *Anal. Chem.* **78**, 8281–8289 (2006).
57. Rogers, D. C. Development of a continuous flow thermal gradient diffusion chamber for ice nucleation studies. *Atmos. Res.* **22**, 149–181 (1988).
58. DeMott, P. J. et al. The Fifth International Workshop on Ice Nucleation phase 2 (FIN-02): laboratory intercomparison of ice nucleation measurements. *Atmos. Meas. Tech.* **11**, 6231–6257 (2018).
59. Schiebel, T. *Ice Nucleation Activity of Soil Dust Aerosols*. PhD thesis, KIT-Bibliothek (2017).
60. Seinfeld, J. H. & Pandis, S. N. *Atmospheric Chemistry and Physics* 2nd edn (Wiley, 2006).
61. Connolly, P. J. et al. Studies of heterogeneous freezing by three different desert dust samples. *Atmos. Chem. Phys.* **9**, 2805–2824 (2009).
62. Koop, T., Luo, B., Tsias, A. & Peter, T. Water activity as the determinant for homogeneous ice nucleation in aqueous solutions. *Nature* **406**, 611–614 (2000).
63. Cziczo, D. J. & Abbatt, J. P. D. Infrared observations of the response of NaCl, MgCl₂, NH₄HSO₄, and NH₄NO₃ aerosols to changes in relative humidity from 298 to 238 K. *J. Phys. Chem. A* **104**, 2038–2047 (2000).

64. Zuberi, B., Bertram, A. K., Koop, T., Molina, L. T. & Molina, M. J. Heterogeneous freezing of aqueous particles induced by crystallized $(\text{NH}_4)_2\text{SO}_4$, ice, and letovicite. *J. Phys. Chem. A* **105**, 6458–6464 (2001).
65. Schlenker, J. C. & Martin, S. T. Crystallization pathways of sulfate–nitrate–ammonium aerosol particles. *J. Phys. Chem. A* **109**, 9980–9985 (2005).
66. Riccobono, F. et al. Oxidation products of biogenic emissions contribute to nucleation of atmospheric particles. *Science* **344**, 717–721 (2014).
67. Möhler, O. & Arnold, F. Gaseous sulfuric acid and sulfur dioxide measurements in the Arctic troposphere and lower stratosphere: implications for hydroxyl radical abundances. *Geophys. Res. Lett.* **19**, 1763–1766 (1992).
68. Williamson, C. J. et al. Large hemispheric difference in ultrafine aerosol concentrations in the lowermost stratosphere at mid and high latitudes. *Atmos. Chem. Phys. Discuss.* 1–44 (2021).
69. Wespes, C. et al. First global distributions of nitric acid in the troposphere and the stratosphere derived from infrared satellite measurements. *J. Geophys. Res. Atmos.* **112**, D13311 (2007).
70. Popp, P. et al. Stratospheric correlation between nitric acid and ozone. *J. Geophys. Res. Atmos.* **114**, D03305 (2009).
71. Hirsikko, A. et al. Atmospheric ions and nucleation: a review of observations. *Atmos. Chem. Phys.* **11**, 767–798 (2011).
72. Jöckel, P. et al. Development cycle 2 of the modular earth submodel system (MESSy2). *Geosci. Model Dev.* **3**, 717–752 (2010).
73. Roeckner, E. et al. Sensitivity of simulated climate to horizontal and vertical resolution in the ECHAM5 atmosphere model. *J. Clim.* **19**, 3771–3791 (2006).

74. Ehrhart, S. et al. Two new submodels for the Modular Earth Submodel System (MESSy): New Aerosol Nucleation (NAN) and small ions (IONS) version 1.0. *Geosci. Model Dev.* **11**, 4987–5001 (2018).
75. Chipperfield, M. New version of the TOMCAT/SLIMCAT off-line chemical transport model: Intercomparison of stratospheric tracer experiments. *Q. J. R. Meteorol. Soc.* **132**, 1179–1203 (2006).
76. Monks, S. A. et al. The TOMCAT global chemical transport model v1.6: description of chemical mechanism and model evaluation. *Geosci. Model Dev.* **10**, 3025–3057 (2017).
77. Mann, G. et al. Description and evaluation of GLOMAP-mode: a modal global aerosol microphysics model for the UKCA composition-climate model. *Geosci. Model Dev.* **3**, 519–551 (2010).
78. Benduhn, F. et al. Size-resolved simulations of the aerosol inorganic composition with the new hybrid dissolution solver HyDiS-1.0: description, evaluation and first global modelling results. *Geosci. Model Dev.* **9**, 3875–3906 (2016).
79. Bardakov, R. et al. A novel framework to study trace gas transport in deep convective clouds. *J. Adv. Model. Earth Syst.* **12**, e2019MS001931 (2020).
80. Bardakov, R., Thornton, J. A., Riipinen, I., Krejci, R. & Ekman, A. M. L. Transport and chemistry of isoprene and its oxidation products in deep convective clouds. *Tellus B Chem. Phys. Meteorol.* **73**, 1–21 (2021).
81. Savre, J., Ekman, A. M. L. & Svensson, G. Technical note: Introduction to MIMICA, a large-eddy simulation solver for cloudy planetary boundary layers. *J. Adv. Model. Earth Syst.* **6**, 630–649 (2014).
82. Bates, T. S., Huebert, B. J., Gras, J. L., Griffiths, F. B. & Durkee, P. A. International Global Atmospheric Chemistry (IGAC) project's first

aerosol characterization experiment (ACE 1): overview. *J. Geophys. Res. Atmos.* **103**, 16297–16318 (1998).

83. Hoell, J. M. et al. Pacific Exploratory Mission in the tropical Pacific: PEM-Tropics A, August-September 1996. *J. Geophys. Res. Atmos.* **104**, 5567–5583 (1999).
84. Erisman, J. W., Sutton, M. A., Galloway, J., Klimont, Z. & Winiwarter, W. How a century of ammonia synthesis changed the world. *Nat. Geosci.* **1**, 636–639 (2008).
85. Warner, J. X. et al. Increased atmospheric ammonia over the world’s major agricultural areas detected from space. *Geophys. Res. Lett.* **44**, 2875–2884 (2017).
86. Twohy, C. H. et al. Deep convection as a source of new particles in the midlatitude upper troposphere. *J. Geophys. Res. Atmos.* **107**, AAC 6-1–AAC 6-10 (2002).
87. Lee, S.-H. et al. Particle formation by ion nucleation in the upper troposphere and lower stratosphere. *Science* **301**, 1886–1889 (2003).
88. Waddicor, D. A. et al. Aerosol observations and growth rates downwind of the anvil of a deep tropical thunderstorm. *Atmos. Chem. Phys.* **12**, 6157–6172 (2012).
89. Wang, S. et al. Active and widespread halogen chemistry in the tropical and subtropical free troposphere. *Proc. Natl Acad. Sci.* **112**, 9281–9286 (2015).
90. He, X.-C. et al. Role of iodine oxoacids in atmospheric aerosol nucleation. *Science* **371**, 589–595 (2021).

Acknowledgements

We thank the European Organization for Nuclear Research (CERN) for supporting CLOUD with important technical and financial resources. This research has received funding from the US National Science Foundation

(nos. AGS-1801574, AGS-1801897, AGS-1602086, AGS-1801329, AGS-2132089 and AGS-1801280), the European Union’s Horizon 2020 programme (Marie Skłodowska-Curie ITN no. 764991 ‘CLOUD-MOTION’), the European Commission, H2020 Research Infrastructures (FORCeS, no. 821205), the European Union’s Horizon 2020 research and innovation programme (Marie Skłodowska-Curie no. 895875 ‘NPF-PANDA’), a European Research Council (ERC) project ATM-GTP contract (no. 742206), an ERC-CoG grant INTEGRATE (no. 867599), the Swiss National Science Foundation (nos. 200021_169090, 200020_172602 and 20FI20_172622), the Academy of Finland ACCC Flagship (no. 337549), the Academy of Finland Academy professorship (no. 302958), the Academy of Finland (nos. 1325656, 316114 and 325647), Russian MegaGrant project ‘Megapolis – heat and pollution island: interdisciplinary hydroclimatic, geochemical and ecological analysis’ (application reference 2020-220-08-5835), Jane and Aatos Erkko Foundation ‘Quantifying carbon sink, CarbonSink+ and their interaction with air quality’ INAR project, Samsung PM2.5 SRP, Prince Albert Foundation ‘the Arena for the gap analysis of the existing Arctic Science Co-Operations (AASCO)’ (no. 2859), the German Federal Ministry of Education and Research (CLOUD-16 project nos. 01LK1601A and 01LK1601C), the Knut and Alice Wallenberg Foundation Wallenberg Academy Fellows project AtmoRemove (no. 2015.0162), the Portuguese Foundation for Science and Technology (no. CERN/FIS-COM/0014/2017) and the Technology Transfer Project N059 of the Karlsruhe Institute of Technology (KIT). The FIGAERO-CIMS was supported by a Major Research Instrumentation (MRI) grant for the US NSF AGS-1531284, as well as the Wallace Research Foundation. The computations by R.Bardakov were performed on resources provided by the Swedish National Infrastructure for Computing (SNIC) at the National Supercomputer Center (NSC). I.R. thanks the Max Planck Society for a sabbatical award. M.W. thanks Siebel Scholars Foundation for financial support.

Author information

Author notes

1. Mingyi Wang

Present address: Division of Chemistry and Chemical Engineering,
California Institute of Technology, Pasadena, CA, USA

Affiliations

1. Center for Atmospheric Particle Studies, Carnegie Mellon University, Pittsburgh, PA, USA

Mingyi Wang, Brandon Lopez, Victoria Hofbauer, Naser G. A. Mahfouz, Roy L. Mauldin, Meredith Schervish, Hamish Gordon & Neil M. Donahue

2. Department of Chemistry, Carnegie Mellon University, Pittsburgh, PA, USA

Mingyi Wang, Victoria Hofbauer, Roy L. Mauldin, Meredith Schervish & Neil M. Donahue

3. Laboratory of Atmospheric Chemistry, Paul Scherrer Institute, Villigen, Switzerland

Mao Xiao, Ruby Marten, Lubna Dada, Houssni Lamkaddam, Dongyu S. Wang, Urs Baltensperger & Imad El-Haddad

4. Institute of Meteorology and Climate Research, Karlsruhe Institute of Technology, Karlsruhe, Germany

Barbara Bertozzi, Pia Bogert, Kristina Höhler, Nsikanabasi Silas Umo, Franziska Vogel, Robert Wagner & Ottmar Möhler

5. Institute for Atmospheric and Environmental Sciences, Goethe University Frankfurt, Frankfurt am Main, Germany

Guillaume Marie, Lucía Caudillo, Manuel Granzin, Martin Heinritzi, Andreas Kürten, Tatjana Müller, Mario Simon, Marcel Zauner-Wieczorek, Joachim Curtius & Jasper Kirkby

6. Institute for Atmospheric and Earth System Research (INAR),
University of Helsinki, Helsinki, Finland

Birte Rörup, Xu-Cheng He, Jiali Shen, Lubna Dada, Rima Baalbaki, Zoé Brasseur, Jonathan Duplissy, Katrianne Lehtipalo, Tuukka Petäjä, Yusheng Wu, Mikko Sipilä, Markku Kulmala & Douglas R. Worsnop

7. Division of Chemistry and Chemical Engineering, California Institute of Technology, Pasadena, CA, USA

Benjamin Schulze & Richard C. Flagan

8. Department of Meteorology, Stockholm University, Stockholm, Sweden

Roman Bardakov & Annica M. L. Ekman

9. Bolin Centre for Climate Research, Stockholm University, Stockholm, Sweden

Roman Bardakov, Annica M. L. Ekman & Ilona Riipinen

10. Institute for Ion Physics and Applied Physics, University of Innsbruck, Innsbruck, Austria

Wiebke Scholz, Bernhard Mentler & Armin Hansel

11. Department of Chemical Engineering, Carnegie Mellon University, Pittsburgh, PA, USA

Brandon Lopez, Hamish Gordon & Neil M. Donahue

12. CERN, the European Organization for Nuclear Research, Geneva, Switzerland

Hanna E. Manninen, Louis-Philippe De Menezes, Roberto Guida, Serge Mathot, Antti Onnela, Stefan K. Weber & Jasper Kirkby

13. CENTRA and Faculdade de Ciências da Universidade de Lisboa,
Campo Grande, Lisbon, Portugal

António Amorim

14. Leibniz Institute for Tropospheric Research, Leipzig, Germany

Farnoush Ataei

15. Helsinki Institute of Physics, University of Helsinki, Helsinki, Finland

Jonathan Duplissy & Markku Kulmala

16. Department of Chemistry & CIRES, University of Colorado Boulder,
Boulder, CO, USA

Henning Finkenzeller & Rainer Volkamer

17. Faculty of Physics, University of Vienna, Vienna, Austria

Loïc Gonzalez Carracedo & Paul M. Winkler

18. Department of Applied Physics, University of Eastern Finland,
Kuopio, Finland

Kimmo Korhonen & Siegfried Schobesberger

19. Aerodyne Research, Inc., Billerica, MA, USA

Jordan E. Krechmer & Douglas R. Worsnop

20. Finnish Meteorological Institute, Helsinki, Finland

Katrianne Lehtipalo, Ana A. Piedehierro & André Welti

21. Atmospheric and Oceanic Sciences, Princeton University, Princeton,
NJ, USA

Naser G. A. Mahfouz

22. P. N. Lebedev Physical Institute of the Russian Academy of Sciences,
Moscow, Russia

Vladimir Makhmutov, Maxim Philippov & Yuri Stozhkov

23. Moscow Institute of Physics and Technology (National Research
University), Moscow, Russia

Vladimir Makhmutov

24. Department of Physics, University of Genoa & INFN, Genoa, Italy

Dario Massabò

25. Department of Atmospheric and Oceanic Sciences, University of
Colorado Boulder, Boulder, CO, USA

Roy L. Mauldin

26. Atmospheric Chemistry Department, Max Planck Institute for
Chemistry, Mainz, Germany

Tatjana Müller, Andrea Pozzer & Jos Lelieveld

27. School of Earth and Environment, University of Leeds, Leeds, UK

Ananth Ranjithkumar

28. Institute Infante Dom Luíz, University of Beira Interior, Covilhã,
Portugal

António Tomé

29. Ionicon Analytik Ges.m.b.H., Innsbruck, Austria

Armin Hansel

30. Joint International Research Laboratory of Atmospheric and Earth
System Sciences, Nanjing University, Nanjing, China

Markku Kulmala

31. Aerosol and Haze Laboratory, Beijing Advanced Innovation Center for Soft Matter Science and Engineering, Beijing University of Chemical Technology, Beijing, China

Markku Kulmala

32. Department of Environmental Science (ACES), Stockholm University, Stockholm, Sweden

Ilona Riipinen

33. Climate and Atmosphere Research Center, The Cyprus Institute, Nicosia, Cyprus

Jos Lelieveld & Theodoros Christoudias

34. Department of Engineering and Public Policy, Carnegie Mellon University, Pittsburgh, PA, USA

Neil M. Donahue

Contributions

M.W., B.B., J.K. and N.M.D. planned the experiments. M.W., B.B., G.M., B.R., B.S., X.-C.H., J.S., W.S., R.M., B.L., H.L., H.E.M., F.A., P.B., Z.B., L.C., L.-P.D.M., J.D., H.F., L.G.C., M.G., R.G., V.H., A.K., K.L., V.M., D.M., S.M., R.L.M., B.M., T.M., A.O., T.P., M.P., A.A.P., A.P., M.Simon, Y.S., A.T., N.S.U., F.V., R.W., D.S.W., S.K.W., A.W., Y.W., M.Z.-W., M.Sipilä, P.M.W., A.H., U.B., M.K., R.C.F., J.C., R.V., I.E.-H., J.K., K.K. O.M., S.S. and N.M.D. prepared the CLOUD facility or measuring instruments. M.W., B.B., G.M., B.R., B.S., X.-C.H., J.S., W.S., R.M., B.L., H.E.M., A.A., L.C., L.G.C., M.G., M.H., V.H., J.E.K., N.G.A.M., D.M., R.L.M., B.M., A.R., M.Schervish, M.Simon, A.T., N.S.U., F.V., D.S.W., S.K.W., A.W., M.Z.-W., P.M.W., J.K. and K.K. collected the data. M.W., M.X., B.B., G.M., B.R., B.S., R.Bardakov, J.S., W.S., L.D., R.Baalbaki, B.L., D.S.W., S.K.W., A.W., I.R., T.C. and N.M.D. analysed the data. M.W.,

M.X., B.B., R.Bardakov, X.-C.H., J.S., W.S., R.M., L.D., R.Baalbaki., B.L., H.L., H.E.M., A.M.L.E., H.F., M.H., K.H., A.K., N.S.U., R.W., A.W., A.H., U.B., M.K., R.C.F., J.C., R.V., I.R., H.G., J.L., I.E.-H., D.R.W., T.C., J.K., O.M., S.S. and N.M.D. contributed to the scientific discussion. M.W., B.B., R.Bardakov, W.S., R.M., B.L., H.L., K.H., A.K., U.B., R.C.F., J.C., R.V., I.R., H.G., J.L., I.E.-H., T.C., J.K., O.M. and N.M.D. wrote the manuscript.

Corresponding author

Correspondence to [Neil M. Donahue](#).

Ethics declarations

Competing interests

The authors declare no competing interests.

Peer review

Peer review information

Nature thanks Bernd Kärcher and the other, 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 Enhancement of HNO₃–NH₃ particle formation by sulfuric acid.](#)

a, Particle number concentrations versus time at mobility diameters >1.7 nm (magenta) and >2.5 nm (green). The solid magenta trace is measured by a PSM_{1.7} and the solid green trace is measured by a CPC_{2.5}. The fixed experimental conditions are about $6.5 \times 10^8 \text{ cm}^{-3}$ NH₃, 223 K and 25% relative humidity. **b**, Particle formation rate versus time at 1.7 nm ($J_{1.7}$), measured by a PSM. **c**, Particle size distribution versus time, measured by an SMPS. **d**, Gas-phase nitric acid and sulfuric acid versus time, measured by an I[−] CIMS and a NO₃[−] CIMS, respectively. We started the experiment by oxidizing NO₂ to produce $1.6 \times 10^9 \text{ cm}^{-3}$ HNO₃ in the presence of about $6.5 \times 10^8 \text{ cm}^{-3}$ ammonia. At time = 0 min, we turned off the high-voltage clearing field to allow the ion concentration to build up to a steady state between GCR production and wall deposition. The presence of ions (GCR condition) induces slow HNO₃–NH₃ nucleation, followed by relatively fast particle growth by nitric acid and ammonia condensation. We thus observe formation of both 1.7-nm and 2.5-nm particles by about one order of

magnitude in about 3.5 h, with a slower approach to steady state because of the longer wall deposition time constant for the larger particles. Then, we increased H_2SO_4 in the chamber from 0 to $4.9 \times 10^6 \text{ cm}^{-3}$ by oxidizing progressively more injected SO_2 after 211 min, with a fixed production rate of nitric acid and injection rate of ammonia. Subsequently, particle concentrations increase by three orders of magnitude within 30 min. The overall systematic scale uncertainties of $\pm 30\%$ on particle formation rate, $-33\% + 50\%$ on sulfuric acid concentration and $\pm 25\%$ on nitric acid concentration are not shown.

[Source data](#)

[Extended Data Fig. 2 Enhancement of \$\text{H}_2\text{SO}_4\text{-HNO}_3\$ nucleation by ammonia.](#)

a, Particle number concentrations versus time at mobility diameters $>1.7 \text{ nm}$ (magenta) and $>2.5 \text{ nm}$ (green). The solid magenta trace is measured by a $\text{PSM}_{1.7}$ and the solid green trace is measured by a $\text{CPC}_{2.5}$. The fixed experimental conditions are 223 K and 25% relative humidity. **b**, Particle formation rate versus time at 1.7 nm ($J_{1.7}$), measured by a PSM. **c**, Particle size distribution versus time, measured by an SMPS. **d**, Gas-phase nitric acid and sulfuric acid versus time, measured by an I^- CIMS and a NO_3^- CIMS, respectively; gas-phase ammonia versus time, calculated with a first-order wall-loss rate. Before the experiment, we cleaned the chamber by rinsing the walls with ultra-pure water, followed by heating to 373 K and flushing at a high rate with humidified synthetic air for 48 h. We started with an almost perfectly clean chamber and only HNO_3 , SO_2 and O_3 vapours present at constant levels. Sulfuric acid starts to appear by means of SO_2 oxidation soon after switching on the UV lights at time = 0 min, building up to a steady state of $5.0 \times 10^6 \text{ cm}^{-3}$ with the wall-loss timescale of about 10 min. Subsequently, we observe slow formation of 1.7-nm particles, yet they do not reach 2.5 nm during the course of a 2-h period with small growth rates and low survival probability. Then, owing to the injection of ammonia from 0 to around $6.5 \times 10^8 \text{ cm}^{-3}$ into the chamber after 80 min, a sharp increase in the rate of particle formation is observed

with a fixed production rate of sulfuric acid and injection rate of nitric acid. The sulfuric acid concentration decreases slightly afterwards, owing to accumulated condensation sink from fast particle growth. The overall systematic scale uncertainties of $\pm 30\%$ on particle formation rate, $-33\%/50\%$ on sulfuric acid concentration and $\pm 25\%$ on nitric acid concentration are not shown.

[Source data](#)

Extended Data Fig. 3 Particle formation rates at 1.7 nm (*J*_{1.7}) versus ammonia concentration at 223 K and 25% relative humidity.

Circles are the CLOUD measurements (the same as those in Fig. 2). The curve represents the model simulations on the basis of known thermodynamics and microphysics, including Kelvin effects, for nucleating clusters.

[Source data](#)

Extended Data Fig. 4 Measurement of the ice nucleation ability of HNO₃–H₂SO₄–NH₃ particles versus sulfate-to-nitrate ratio.

a, Particle size distribution versus time during the experiment, measured by an SMPS. **b**, Gas-phase sulfuric acid versus time, measured by a nitrate CIMS. **c**, Particle-phase chemical composition versus time, measured by an AMS. **d**, Fraction of INP at the nominal temperature of 215 K. The horizontal black dashes indicate the ice fraction threshold, $f_{\text{ice}} = 10^{-3}$. The coloured circles correspond to the sulfate-to-nitrate ratios shown in Fig. 4a.

[Source data](#)

Extended Data Fig. 5 Parameterization of the HNO₃–H₂SO₄–NH₃ particle formation rate.

a–c Particle formation rate ($J_{1.7}$) as a function of H_2SO_4 , HNO_3 and NH_3 vapour concentrations, respectively, at 223 K and 25% relative humidity. The red triangles, blue circles and yellow squares represent experiments while varying only the concentration of H_2SO_4 (Extended Data Fig. 1), HNO_3 (Fig. 1) and NH_3 (Extended Data Fig. 2), respectively. The H_2SO_4 concentration was varied between 4.6×10^5 and $2.9 \times 10^6 \text{ cm}^{-3}$, HNO_3 between 2.3×10^8 and $1.7 \times 10^9 \text{ cm}^{-3}$ and NH_3 between 1.8×10^8 and $5.1 \times 10^8 \text{ cm}^{-3}$. **d**, The multi-acid–ammonia parameterization (black line) on the basis of equation (6) with $k = 3.4 \times 10^{-71} \text{ s}^{-1} \text{ cm}^{24}$. The grey dashed horizontal line shows a maximum of about $2 \text{ cm}^{-3} \text{ s}^{-1}$ ion-induced nucleation in the CLOUD chamber under GCR conditions, limited by the ion-pair production rate from GCR plus beam-background muons. The bars indicate 30% estimated total error on the particle formation rates, although the overall systematic scale uncertainties of $-33\% / +50\%$ on sulfuric acid concentration and $\pm 25\%$ on nitric acid concentration are not shown. **e**, Temperature dependence of $J_{1.7}$ for $\text{HNO}_3\text{--H}_2\text{SO}_4\text{--NH}_3$ nucleation (blue curve) on the basis of equation (9) with $k = 2.9 \times 10^{-98} e^{14,000/T} \text{ s}^{-1} \text{ cm}^{24}$.

[Source data](#)

[Extended Data Fig. 6 Modelled contribution of \$\text{HNO}_3\text{--H}_2\text{SO}_4\text{--NH}_3\$ nucleation to upper tropospheric particles.](#)

Number concentrations of multi-acid new particles (nucleation mode) at 250-hPa altitude simulated in a global model (EMAC) with efficient vertical transport of ammonia. The particle formation rate is on the basis of the blue dashed curve in Fig. 2 and parameterization shown in Extended Data Fig. 5. The extra particle number concentrations are shown, that is, relative to the same model without multi-acid nucleation. High annually averaged particle numbers are expected in the monsoon region (grey rectangle) and adjacent regions.

[Extended Data Fig. 7 Modelled annual mean ammonia mixing ratios at 250 hPa \(11 km, about 223 K\).](#)

a, The EMAC global model simulations are higher than the MIPAS satellite observations, although consistent with aircraft measurements^{5,6}. **b**, The TOMCAT global model predicts much less ammonia (<1 pptv) in the upper troposphere.

Extended Data Fig. 8 Modelled transport of ammonia to the upper troposphere in deep convective clouds.

a, Trajectories of the simulated convective cloud event (grey) and a selected parcel representing a buoyant parcel reaching the upper troposphere (black). **b**, The simulated evolution of parcel A altitude (green dashed trace) and the total mass concentration and phase of the cloud hydrometeors (red and blue curves). **c–e** Sensitivity of the predicted ammonia concentrations within parcel A to cloud water pH, total water amount and retention coefficient (by ice particles) as compared with the base-case simulation (blue trace in all figures).

[Source data](#)

Supplementary information

[Peer Review File](#)

Source data

[Source Data Fig. 1](#)

[Source Data Fig. 2](#)

[Source Data Fig. 3](#)

[Source Data Fig. 4](#)

[Source Data Extended Data Fig. 1](#)

[Source Data Extended Data Fig. 2](#)

[Source Data Extended Data Fig. 3](#)

[Source Data Extended Data Fig. 4](#)

[Source Data Extended Data Fig. 5](#)

[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

Wang, M., Xiao, M., Bertozzi, B. *et al.* Synergistic HNO₃–H₂SO₄–NH₃ upper tropospheric particle formation. *Nature* **605**, 483–489 (2022). <https://doi.org/10.1038/s41586-022-04605-4>

- Received: 28 July 2021
- Accepted: 02 March 2022
- Published: 18 May 2022
- Issue Date: 19 May 2022
- DOI: <https://doi.org/10.1038/s41586-022-04605-4>

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/s41586-022-04605-4>

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

[Download PDF](#)

- Article
- Open Access
- [Published: 11 May 2022](#)

Genetic and chemotherapeutic influences on germline hypermutation

- [Joanna Kaplani¹](#),
- [Benjamin Ide²](#),
- [Rashesh Sanghvi](#) [ORCID: orcid.org/0000-0002-7703-9216¹](#),
- [Matthew Neville](#) [ORCID: orcid.org/0000-0001-5816-7936¹](#),
- [Petr Danecek¹](#),
- [Tim Coorens](#) [ORCID: orcid.org/0000-0002-5826-3554¹](#),
- [Elena Prigmore](#) [ORCID: orcid.org/0000-0001-8870-0316¹](#),
- [Patrick Short¹](#),
- [Giuseppe Gallone¹](#),
- [Jeremy McRae](#) [ORCID: orcid.org/0000-0003-3411-9248¹](#),
- [Genomics England Research Consortium](#),
- [Jenny Carmichael³](#),
- [Angela Barnicoat⁴](#),
- [Helen Firth^{1,3}](#),
- [Patrick O'Brien](#) [ORCID: orcid.org/0000-0001-7853-8626²](#),
- [Raheleh Rahbari](#) [ORCID: orcid.org/0000-0002-1839-7785¹](#) &
- [Matthew Hurles](#) [ORCID: orcid.org/0000-0002-2333-7015¹](#)

[Nature](#) volume 605, pages 503–508 (2022)

- 7663 Accesses

- 170 Altmetric
- [Metrics details](#)

Subjects

- [Genomics](#)
- [Medical genetics](#)
- [Mutation](#)
- [Neurodevelopmental disorders](#)

Abstract

Mutations in the germline generates all evolutionary genetic variation and is a cause of genetic disease. Parental age is the primary determinant of the number of new germline mutations in an individual's genome^{1,2}. Here we analysed the genome-wide sequences of 21,879 families with rare genetic diseases and identified 12 individuals with a hypermutated genome with between two and seven times more de novo single-nucleotide variants than expected. In most families (9 out of 12), the excess mutations came from the father. Two families had genetic drivers of germline hypermutation, with fathers carrying damaging genetic variation in DNA-repair genes. For five of the families, paternal exposure to chemotherapeutic agents before conception was probably a key driver of hypermutation. Our results suggest that the germline is well protected from mutagenic effects, hypermutation is rare, the number of excess mutations is relatively modest and most individuals with a hypermutated genome will not have a genetic disease.

Main

The average number of de novo mutations (DNMs) generating single-nucleotide variants (SNVs) is estimated to be 60–70 per human genome per generation, but little is known about individuals with germline hypermutation with unusually large numbers of DNMs^{1,3,4}. The human germline-mutation rate varies between individuals, families and populations, and has evolved over time^{5,6,7,8,9}. Parental age explains a large proportion of

variance for SNVs, insertion–deletions (indels) and short tandem repeats^{1,10,11}. It has been estimated that there is an increase of around 2 DNM s for every additional year in father’s age and around 0.5 DNM s for every additional year in mother’s age^{1,12}. Subtle differences have also been observed between the maternal and paternal mutational spectra and may be indicative of different mutagenic processes^{2,13,14,15}. Different mutational mechanisms can leave distinct mutational patterns termed ‘mutational signatures’^{16,17}. There are currently more than 100 somatic mutational signatures that have been identified across a wide variety of cancers of which half have been attributed to endogenous mutagenic processes or specific mutagens^{18,19}. The majority of germline mutations can be explained by two of these signatures, termed signature 1 (SBS1), probably due to deamination of 5-methylcytosine²⁰, and signature 5 (SBS5), which is thought to be a pervasive and relatively clock-like endogenous process. Both signatures are ubiquitous among normal and cancer cell types^{21,22} and have been reported previously in trio studies¹⁴. The impact of environmental mutagens has been well established in the soma but is not as well understood in the germline^{23,24}. Environmental exposures in parents, such as ionizing radiation, can influence the number of mutations transmitted to offspring^{25,26,27}. Individual mutation rates can also be influenced by genetic background. With regard to somatic mutation, thousands of inherited germline variants have been shown to increase cancer risk^{28,29,30}. Many of these variants are in genes that encode components of DNA-repair pathways which, when impaired, lead to an increase in the number of somatic mutations. However, it is unclear whether variants in known somatic mutator genes can influence germline-mutation rates. There are examples in which the genetic background has been shown to affect the local germline-mutation rate of short tandem repeats, minisatellites and translocations^{31,32,33,34,35}.

An increasing germline-mutation rate results in an increased risk of offspring being born with a dominant genetic disorder³⁶. Long-term effects of mutation rate differences as a result of mutation accumulation have been demonstrated in mice to have effects on reproduction and survival rates and there may be a similar impact in humans^{37,38}.

Little is known about rare outliers with extreme mutation rates. DNMs are a substantial cause of rare genetic disorders and cohorts of patients with such disorders are more likely to include individuals with germline hypermutation [12,39](#). To this end, we sought to identify individuals with germline hypermutation in sequenced parent–offspring trios from two rare disease cohorts. We identified genetic or environmental causes of this hypermutation and estimated how much variation in the germline-mutation rate that this may explain.

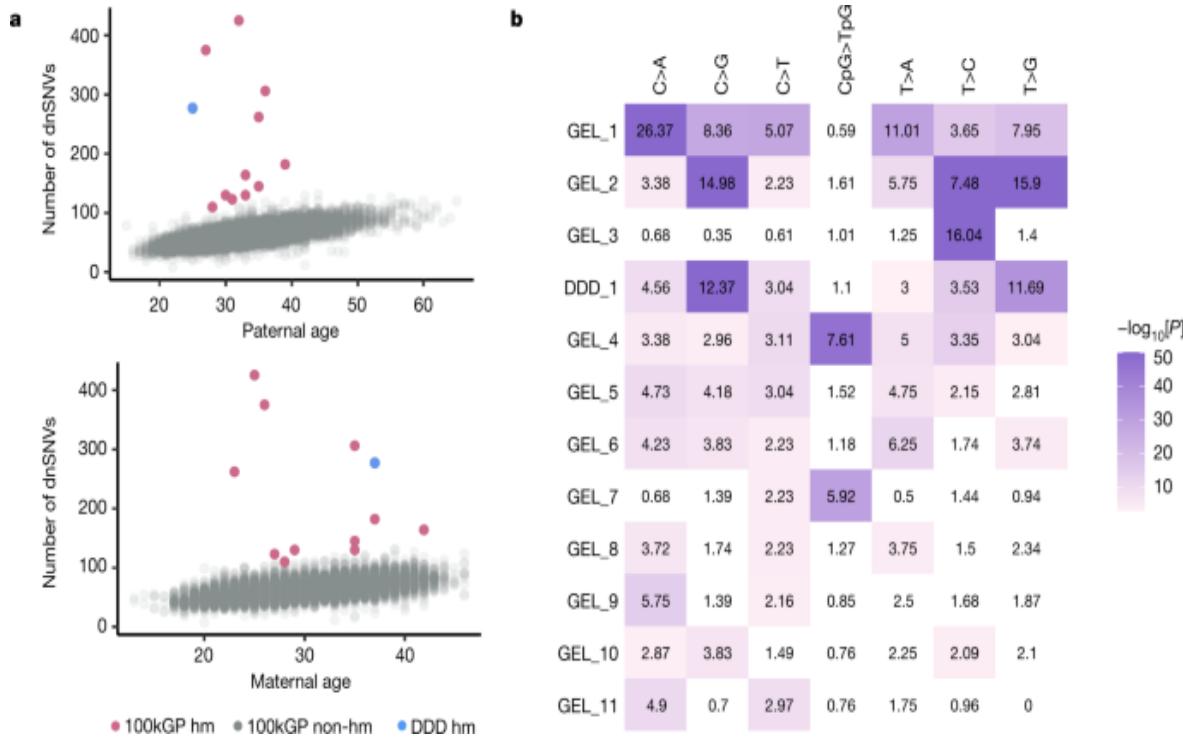
Individuals with germline hypermutation

We identified individuals with germline hypermutation in two separate cohorts comprising parent–offspring trios: 7,930 exome-sequenced trios from the Deciphering Developmental Disorders (DDD) study and 13,949 whole-genome sequenced trios in the rare disease arm of the 100,000 Genome Project (100kGP). We selected nine trios from the DDD study with the largest number of DNMs, given their parental ages, which were subsequently whole-genome sequenced to characterize DNMs genome-wide. In the 100kGP cohort, we performed filtering of the DNMs, which resulted in a total of 903,525 de novo SNVs (dnSNVs) and 72,110 de novo indels (dnIndels). The median number of DNMs per individual was 62 for dnSNVs and 5 for dnIndels (median paternal and maternal ages of 33 and 30) (Supplementary Fig. 1).

We observed an increase in the total number of dnSNVs of 1.28 dnSNVs per year of paternal age (95% confidence interval (CI) = 1.24–1.32, $P < 10^{-300}$, negative binomial regression) and an increase of 0.35 dnSNVs per year of maternal age (95% CI = 0.30–0.39, $P = 3.0 \times 10^{-49}$, negative binomial regression) (Fig. [1a](#)). We phased 241,063 dnSNVs and found that 77% were paternal in origin, in accordance with previous estimates [13,14,15](#). Estimates of the parental age effect in the phased mutations were similar to the unphased results: 1.23 paternal dnSNVs per year of paternal age (95% CI = 1.14–1.32, $P = 1.6 \times 10^{-158}$) and 0.38 maternal dnSNVs per year of maternal age (95% CI = 0.35–0.41, $P = 6.6 \times 10^{-120}$) (Extended Data Fig. [1](#)). Paternal and maternal age were also significantly associated with the number of dnIndels: an increase of 0.071 dnIndels per year of paternal age (95% CI = 0.062–

0.080, $P = 8.3 \times 10^{-56}$; Extended Data Fig. 1) and a smaller increase of 0.019 dnIndels per year of maternal age (95% CI = 0.0085–0.029, $P = 3.4 \times 10^{-4}$; Extended Data Fig. 1). The ratios of paternal to maternal mutation increases per year were very similar—3.7 for SNVs and 3.8 for indels. The proportion of DNMs that phase paternally increased by 0.0017 for every year of paternal age ($P = 3.37 \times 10^{-38}$, binomial regression; Supplementary Fig. 2). However, the proportion of DNMs that phase paternally in the youngest fathers remains around 0.75 and, therefore, the paternal age effect alone does not fully explain the strong paternal bias¹⁵. We compared the mutational spectra of the phased DNMs and found that maternally derived DNMs have a significantly higher proportion of C>T mutations (0.27 maternal versus 0.22 paternal, $P = 3.24 \times 10^{-80}$, binomial test), whereas paternally derived DNMs have a significantly higher proportion of C>A, T>G and T>C mutations (C>A: 0.08 maternal versus 0.10 paternal, $P = 4.6 \times 10^{-23}$; T>G: 0.06 versus 0.7, $P = 6.8 \times 10^{-28}$; T>C: 0.25 versus 0.26, $P = 1.6 \times 10^{-5}$, binomial test; Extended Data Fig. 2a). These mostly agree with previous studies, although the difference in T>C mutations was not previously significant¹³. Most paternal and maternal mutations could be explained by SBS1 and SBS5, with a slightly higher contribution of SBS1 in paternal mutations (0.16 paternal versus 0.15 maternal, χ^2 test, $P = 2.0 \times 10^{-5}$; Extended Data Fig. 2b).

Fig. 1: Identification of individuals with germline hypermutation.

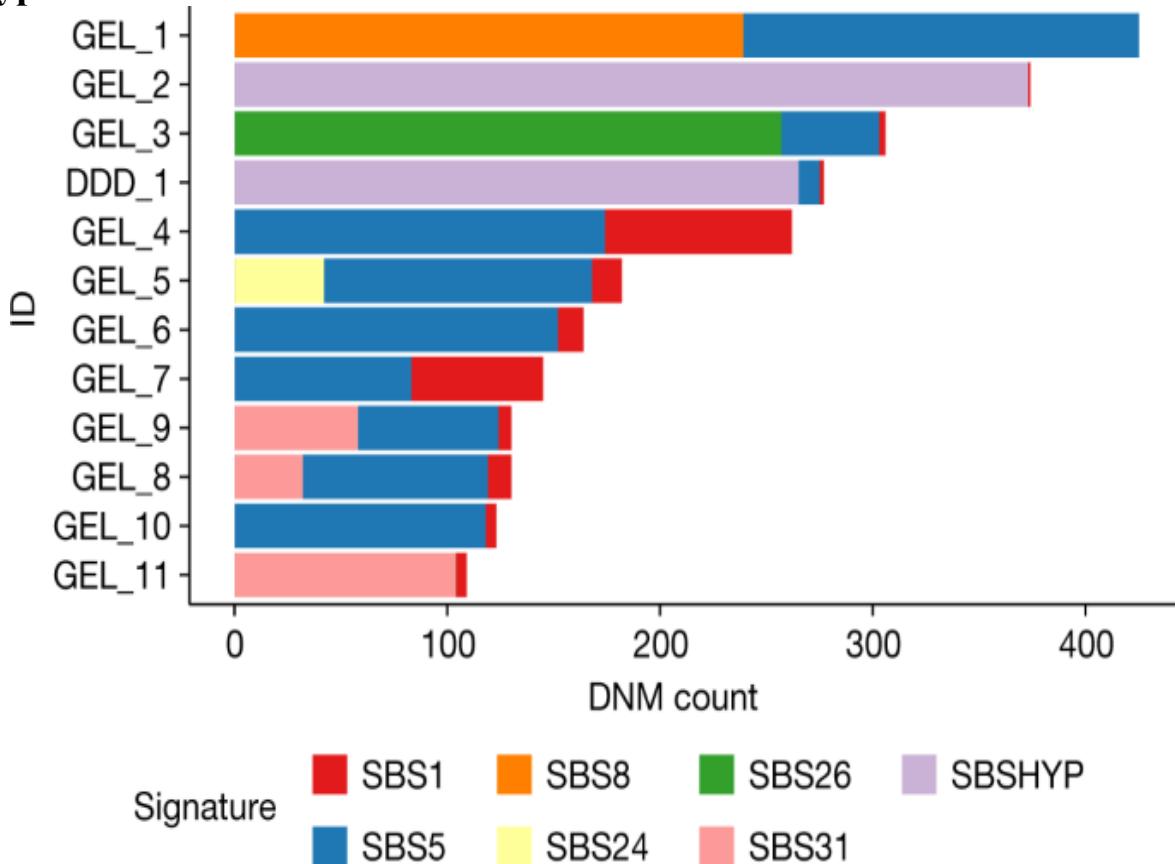


a, Paternal and maternal age versus the number of dnSNVs. Individuals with hypermutation (hm) from the 100kGP cohort (pink) and individuals with hypermutation from the DDD cohort (blue) are highlighted. **b**, Enrichment (observed/expected) of mutation type for individuals with hypermutation. Sample names are shown on the *y* axis, and mutation type is shown on the *x* axis. The enrichment is coloured by the $-\log_{10}[\text{enrichment } P \text{ value}]$, determined using two-sided Poisson tests comparing the average number of mutations in each type across all individuals in the 100kGP cohort. White colouring indicates no statistically significant enrichment after multiple-testing correction ($P < 0.05/12 \times 7$ tests). Exact *P* values are provided in Supplementary Table 2.

We identified 12 individuals with germline hypermutation after accounting for parental age ([Methods](#)): 11 from 100kGP and 1 from DDD (Fig. 1a and Extended Data Table 1). The number of dnSNVs for each of the 12 individuals with hypermutation ranged from 110 to 425, corresponding to a fold increase of 1.7–6.5 compared with the median number of dnSNVs per individual. Two of these individuals also had a significantly increased number of dnIndels (Extended Data Table 1). The mutational spectra across these individuals with hypermutation varied considerably (Fig. 1b, Extended Data Figs. 3 and 4 and Supplementary Tables 1 and 2) and, after extracting

mutational signatures, we found that, although most mutations mapped onto known somatic signatures from COSMIC⁴⁰, a new signature, SBSHYP, was also extracted (Fig. 2, Extended Data Fig. 5 and Supplementary Table 3). In addition to mutational spectra, we evaluated the parental phase, transcriptional strand bias (Extended Data Fig. 6) and the distribution of the variant allele fraction (VAF) for these mutations (Extended Data Fig. 7). After examining these properties, we identified three potential sources of germline hypermutation: paternal defects in DNA-repair genes, paternal exposure to chemotherapeutics and post-zygotic mutational factors.

Fig. 2: Mutational signatures in individuals with germline hypermutation.



Contributions of mutational signatures extracted using SigProfiler and decomposed onto known somatic mutational signatures as well as the signature SBSHYP that we identified in both DDD_1 and GEL_2.

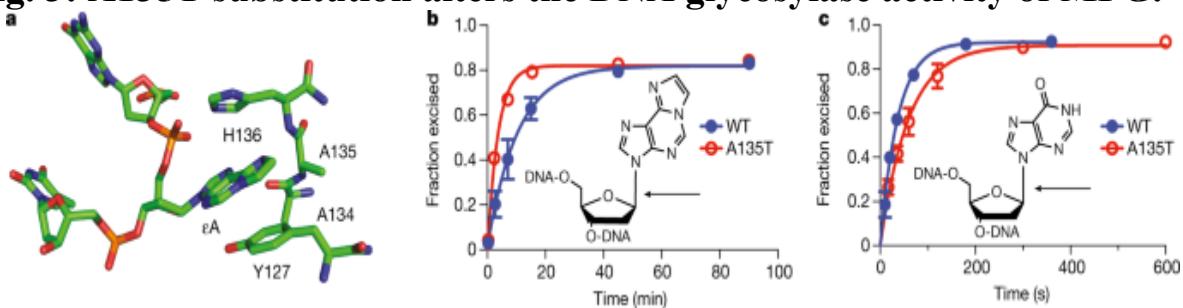
Paternal defects in DNA-repair genes

For eight individuals with hypermutation, the DNMs phased paternally significantly more than expected ($P < 0.05/12$ tests, two-sided binomial test; Extended Data Table 1), implicating the paternal germline as the origin of the hypermutation. Two of these fathers carry rare homozygous non-synonymous variants in known DNA-repair genes (Supplementary Table 4). Defects in DNA repair are known to increase the mutation rate in the soma and may have a similar effect in the germline. Individual GEL_1 has the highest number of DNMs of all of the individuals, and a significantly increased number of dnIndels. The mutational spectra exhibit enrichment of C>A and T>A mutations (Fig. 1b) and we observed a large contribution of the signature SBS8 (Fig. 2). This signature is associated with transcription-coupled nucleotide-excision repair (NER) and typically presents with transcriptional strand bias. This agrees with the strong transcriptional strand bias observed in GEL_1 ($P = 2.1 \times 10^{-40}$, Poisson test; Extended Data Fig. 6). The father has a rare homozygous nonsense variant in the gene *XPC* (Extended Data Table 1 and Supplementary Table 4), which is involved in the early stages of the NER pathway. The paternal variant is annotated as pathogenic for xeroderma pigmentosum in ClinVar and the father had already been diagnosed with this disorder. Patients with xeroderma pigmentosum have a high risk of developing skin cancer and have an increased risk of developing other cancers^{41,42}. *XPC* deficiency has been associated with a similar mutational spectrum to the one that we observed in GEL_1 (ref. 43) and *XPC* deficiency in mice has been shown to increase the germline-mutation rate at two short tandem repeat loci⁴⁴.

GEL_3 has about a fivefold enrichment of dnSNVs, which exhibit a distinctive mutational spectrum with around a seventeenfold increase in T>C mutations but no increase in other mutations (Fig. 1b and Extended Data Fig. 3d). Extraction of mutational signatures revealed that the majority of mutations mapped onto SBS26, which has been associated with defective mismatch repair. The father has a rare homozygous missense variant in the gene *MPG* (Extended Data Table 1 and Supplementary Table 5). *MPG* encodes *N*-methylpurine DNA glycosylase (also known as alkyladenine-DNA glycosylase), which is involved in the recognition of base lesions, including alkylated and deaminated purines, and initiation of the base-excision repair pathway. The *MPG* variant is rare in gnomAD (allele frequency = 9.8×10^{-5} , no observed homozygotes) and is predicted to be

pathogenic (CADD score = 27.9) and the amino acid residue is fully conserved across 172 aligned protein sequences from VarSite^{45,46}. The variant amino acid forms part of the substrate-binding pocket and probably affects substrate specificity (Fig. 3a). *MPG* has not yet been described as a cancer-susceptibility gene, but studies in yeast and mice have demonstrated variants in this gene and, specifically, in its substrate-binding pocket, can lead to a mutator phenotype^{47,48} (Supplementary Table 6). We examined the functional impact of the observed A135T variant using in vitro assays (Methods and Extended Data Figs. 8 and 9). The A135T variant caused a twofold decrease in excision efficiency of the deamination product hypoxanthine (Hx) in both the T and C contexts (Fig. 3c and Extended Data Fig. 9), with a small increase in excision efficiency of an alkylated adduct 1,N(6)-ethenoadenine (ϵ A) in both the T and C contexts (Fig. 3b and Extended Data Fig. 8). The maximal rate of excision is increased by twofold for ϵ A—among the largest increases that have been observed for 15 reported *MPG* variants (Supplementary Table 5). Another variant—N169S, which also shows an increase in *N*-glycosidic bond cleavage with the ϵ A substrate—has been established as a mutator in yeast^{48,49}. These assays confirm that the A135T substitution alters the *MPG*-binding pocket and changes the activity towards different DNA adducts. *MPG* acts on a wide variety of DNA adducts and further functional characterization and mechanistic studies are required to link the observed T>C germline mutational signature to the aberrant processing of a specific class of DNA adducts.

Fig. 3: A135T substitution alters the DNA glycosylase activity of MPG.



a, Active-site view of MPG bound to ϵ A-DNA from Protein Data Bank 1EWN. Ala135 and His136 form the binding pocket for the flipped-out base lesion, which is bracketed by Tyr127 on the opposing face. **b**, Single-turnover excision of ϵ A from ϵ A•T is twofold faster for A135T (red) than for wild-type (blue) MPG. **c**, Single-turnover excision of Hx from Hx•T is

slower for A135T (red) compared with wild-type (blue) MPG. The arrows indicate the *N*-glycosidic bond that is cleaved by MPG. Data are mean \pm s.d. for glycosylase reactions with 10 nM DNA substrate and either 100 nM enzyme for ϵ A excision ($n = 6$) or 40 nM enzyme for Hx excision ($n = 3$) (Extended Data Fig. 9).

Parental chemotherapy before conception

Three individuals with hypermutation (GEL_8, GEL_9 and GEL_11) have a contribution from the signature SBS31 (Fig. 2), which has been associated with treatment with platinum-based drugs, which damage DNA by causing covalent adducts¹⁶. The phased dnSNVs in GEL_9 and GEL_11 are paternally biased (46 paternal:2 maternal, $P = 0.0014$; 28 paternal:1 maternal, $P = 0.012$; binomial test; Extended Data Table 1), and the dnSNVs in GEL_11, who has the largest contribution of SBS31, exhibit a significant transcriptional strand bias, as expected for this signature ($P = 6.9 \times 10^{-6}$, two-sided Poisson test; Extended Data Table 1 and Extended Data Fig. 6). All three fathers had a cancer diagnosis and chemotherapy treatment before conception of their child with a hypermutated genome. The father of GEL_11 was diagnosed with and received chemotherapeutic treatment for osteosarcoma, lung cancer and cancer of the intestinal tract before conception. Cisplatin is a commonly used chemotherapeutic agent for osteosarcoma and lung cancer. Cisplatin mainly reacts with purine bases, forming intrastrand cross-links that can be repaired by NER or bypassed by translesion synthesis, which may in turn induce single-base substitutions⁵⁰. The fathers of GEL_8 and GEL_9 both have a history of testicular cancer where cisplatin is the most commonly administered chemotherapeutic.

GEL_2 and DDD_1 have a similar number of dnSNVs, which are significantly paternally biased (Extended Data Table 1), and share a mutational signature (SBSHYP) that is characterized by an enrichment of C>G and T>G mutations (Fig. 2 and Extended Data Fig. 5) and does not map on to any previously described signatures observed in COSMIC or in response to mutagenic exposure^{24,40,51,52} (Supplementary Fig. 3a). The fathers do not share rare non-synonymous variants in any genes. Both fathers received chemotherapy treatment before conception, including nitrogen mustard alkylating agents (Supplementary Table 5), although with

different members of this class of chemotherapies. We therefore strongly suspect that this class of chemotherapeutic agents is the cause of this mutational signature. Experimental studies of a subset of alkylating agents have shown them to have diverse mutational signatures^{24,51,52,53} (Supplementary Fig. 3b).

GEL_5 has 182 dnSNVs and a significant paternal bias in the phased dnSNVs ($P = 5.8 \times 10^{-4}$, binomial test; Extended Data Table 1). The father of GEL_5 was diagnosed with systemic lupus erythematosus and received chemotherapy treatment before conception; however, the dnSNVs do not map onto any known chemotherapeutic mutational signatures (Figs. 1b and 2). GEL_5 has a contribution of SBS24, which is associated with aflatoxin exposure in cancer blood samples²²; however, aflatoxin exposure is often dietary and there is no evidence of exposure in the father's hospital records

We assessed how parental cancer and exposure to chemotherapy might impact the germline-mutation rate more generally by examining 100kGP hospital records for ICD10 codes related to cancer and chemotherapy recorded before the conception of the child. We identified 27 fathers (0.9%) who had a history of cancer, 7 of whom had testicular cancer (Supplementary Table 7). The offspring of these 27 fathers did not have a significantly increased number of dnSNVs after correcting for parental age ($P = 0.73$, two-sided Wilcoxon test) and their fathers were not significantly older than average ($P = 0.77$, two-sided Wilcoxon test). The available health records are not definitive with regard to historical chemotherapeutic treatments or the potential use of sperm stored before treatment for conception (only 6 had chemotherapy-related ICD10 codes). Although the total number of dnSNVs is not significantly increased, 2 out of the 27 fathers had children with a hypermutated genome, a significant enrichment compared with fathers without a history of cancer (2 out of 27 versus 9 out of 2,891, $P = 0.0043$, Fisher exact test). This is probably a conservative assessment as two other individuals with hypermutation have fathers who were subsequently shown to have had chemotherapy treatment but were not included in this analysis as they did not have any ICD10 codes recorded before conception (Methods). We performed the same analysis across 5,508 mothers in the 100kGP cohort with hospital records before conception and identified 27 mothers (0.5%) with a history of cancer, 9 of whom had

chemotherapy recorded. Children of these 27 mothers had a nominally significant increase in dnSNVs after correcting for parental age and data quality ($P = 0.03$, two-sided Wilcoxon test). These mothers were significantly older at the birth of the child compared with the mothers without cancer ($P = 0.003$, Wilcoxon test). Matching on parental age, children of mothers with cancer had a median increase of 9 dnSNVs.

Among the offspring who did not have a hypermutated genome but had a parental history of cancer, we found only one with unusual mutational signatures (Supplementary Fig. 4). PatCancer_10 has 94 dnSNVs ($P = 0.005$, dnSNV P value after correcting for parental age) of which 89% phased paternally (Supplementary Fig. 4 and Supplementary Table 2) with a contribution from SBS31, which is associated with platinum-based drugs (Supplementary Fig. 4). Their father was treated for testicular cancer before conception.

Post-zygotic hypermutation

Two individuals with hypermutation, GEL_4 and GEL_7, had around a fourfold and twofold increase in dnSNVs, respectively, that phase equally between maternal and paternal chromosomes. The VAF of dnSNVs in these individuals was shifted below 0.5 (Extended Data Fig. 7): the proportion of dnSNVs with $\text{VAF} < 0.4$ was significantly higher compared with all dnSNVs observed (GEL_4: $P = 3.9 \times 10^{-59}$; GEL_7: $P = 8.3 \times 10^{-4}$, two-sided binomial test). These mutations most likely occurred post-zygotically and are not due to a parental hypermutator. Both individuals share a large contribution from SBS1 (ref. 40) (Fig. 2). GEL_4 has several blood-related clinical phenotypes, including myelodysplasia. The observations in GEL_4 are probably due to clonal haematopoiesis leading to a large number of somatic mutations in the child's blood. We identified a mosaic de novo missense mutation in *ETV6*, a gene that is associated with leukaemia and thrombocytopaenia⁵⁴. We did not observe similar blood-related phenotypes in GEL_7 (although the child was one year old at recruitment), nor did we identify a possible genetic driver of clonal haematopoiesis. We investigated whether a maternal protein with a mutator variant may be affecting the mutation rate in the first few cell divisions. We identified a mosaic maternal missense variant in *TP53* that was previously annotated as pathogenic for

Li–Fraumeni cancer predisposition syndrome, which was not observed in the child. It is not known whether this variant is present in the maternal germline or whether it would have a germline mutagenic effect⁵⁵.

Variation in the germline-mutation rate

We investigated the factors influencing the number of dnSNVs per individual in a subset of 7,700 100kGP trios filtered more stringently for data quality ([Methods](#)). We estimated that parental age accounts for 69.7% and data quality metrics explain 1.3% of the variance. The variance explained by parental age is smaller than a previous estimate of 95% on the basis of a sample of 78 families¹. Repeated estimates of the variance explained by parental age from resampling of 78 trios from 100kGP showed that these estimates can vary widely (median = 79%, 95% CI = 52–100%); 7% of resamplings have an estimated variance explained of 95% or greater. We estimated that germline hypermutation in the 11 100kGP individuals with hypermutation explained an additional 7.1% of variance in this cohort. This leaves 21.9% (19.7–23.8%, bootstrap 95% CI) of variance in the numbers of dnSNVs per individual unaccounted for.

Both mutagenic exposures and genetic variation in DNA-repair genes could have a more subtle role in influencing variation in the germline-mutation rate. Moreover, polygenic effects and gene by environment interactions may also contribute. We investigated whether rare variants in DNA-repair genes influence germline-mutation rates in the 100kGP cohort. We curated three sets of rare non-synonymous variants with increasing likelihoods of impacting the germline-mutation rate: (1) variants in all DNA-repair genes ($n = 186$), (2) variants in DNA-repair genes that are most likely to create SNVs ($n = 66$) and (3) the subset of (2) that has been associated with cancer ([Methods](#)). We focused on heterozygous variants (MAF < 0.001), but also considered rare homozygous variants (MAF < 0.01) in all DNA-repair genes. There was no statistically significant effect in any of these groups of variants after Bonferroni correction (Supplementary Fig. 5 and Supplementary Table 8). We examined heterozygous protein-truncating variants (PTVs) in the known cancer mutator gene *MBD4* that are associated with a threefold increased CpG>TpG mutation rate in tumours. We performed whole-genome sequencing of 13 DDD trios with paternal carriers

of *MBD4* PTVs. We found no significant increase in either the total number of DNMs or the number of CpG>TpG mutations ($P = 0.56$, χ^2 ; Supplementary Fig. 6). Power modelling suggested that there is probably not an increase in the CpG germline-mutation rate of higher than a 22%.

To examine potential polygenic contributions, we estimated the residual variation in the number of dnSNVs in the 100kGP cohort (after correcting for parental age, data quality and hypermutation status) explained by more common genetic variants. We estimated this separately for fathers and mothers using GREML-LDMS⁵⁶ stratified by minor allele frequency and linkage disequilibrium. We found that maternal germline variation ($\text{MAF} > 0.001$) is unlikely to explain much residual variation ($h^2 = 0.07$, $P = 0.21$, GCTA reported results; Supplementary Table 9). We found that paternal variation could contribute a substantial fraction of residual variation ($h^2 = 0.53$, 95% CI = 0.20–0.85, $P = 0.09$); however, this seems to be concentrated exclusively in low-frequency variants ($0.001 < \text{MAF} < 0.01$, $h^2 = 0.52$, 95% CI = 0.01–0.94) rather than more common variants ($\text{MAF} > 0.01$, $h^2 = 0.008$, 95% CI = 0–0.38; Supplementary Table 9). Further investigation of polygenic contributions will require larger sample sizes.

Discussion

Germline hypermutation is an uncommon but important phenomenon. We identified 12 individuals with hypermutation from over 20,000 parent–offspring sequenced trios in the DDD and 100kGP cohorts with a two- to sevenfold increase in the number of dnSNVs. There are probably other individuals with germline hypermutation in the DDD cohort, as screening this exome-sequenced cohort for potential individuals with hypermutation for confirmation by genome sequencing will have missed some individuals with two- to sevenfold hypermutation.

In two individuals with hypermutation, the excess mutations occurred post-zygotically; however, for the majority ($n = 8$), excess dnSNVs phased paternally, implicating the father as the source of hypermutation. For five of these fathers, mutational signatures and clinical records implicated the mutagenicity of two classes of chemotherapeutics: platinum-based drugs ($n = 3$) and mustard-derived alkylating agents ($n = 2$). For two fathers,

functional and clinical data implicated the mutagenicity of homozygous missense variants in the known DNA-repair genes *XPC* and *MPG*.

Our findings imply that defects in DNA-repair genes can increase germline-mutation rates in addition to their well-established impacts on somatic mutation rates⁵⁷. However, DNA-repair defects do not always behave similarly in the soma and the germline. We found that PTVs in an established somatic mutator gene, *MBD4*, did not have a detectable effect in the germline⁵⁸. We also did not observe a significant effect on germline-mutation rates of rare non-synonymous variants in DNA-repair genes more generally. Paternal variants previously associated with cancer had a nominally significant effect but amounted to an average increase of only around 2 dnSNVs. Both larger sample sizes and additional variant curation will probably be needed to investigate this further. Genes and pathways that impact germline mutation more than the soma may also exist; detecting mutagenic variants in these genes will be challenging.

Germline hypermutation accounted for 7% of the variance in the germline-mutation rate in the 100kGP cohort. The ascertainment in this cohort for rare genetic diseases probably means that individuals with germline hypermutation are enriched relative to the general population. As a consequence, our estimate of the contribution of germline hypermutation is probably inflated. However, the absolute risk of an individual with a hypermutated germline having a child with a genetic disease is low. The population average risk for having a child with a severe developmental disorder caused by a DNM has been estimated to be 1 in 300 births¹² and so a fourfold increase in DNMs in a child would increase this absolute risk to just over 1%. Thus, most individuals with germline hypermutation will not have a genetic disease, and germline hypermutation should also be observed in healthy individuals.

The two genetic causes of germline hypermutation that we identified were both recessive in action. Similarly, most DNA-repair disorders act recessively in their cellular mutagenic effects. This implies that genetic causes of germline hypermutation are likely to arise at substantially higher frequencies in populations with high rates of parental consanguinity. In such populations, the overall incidence of germline hypermutation may be higher, and the proportion of variance in the number of dnSNVs per offspring

accounting for genetic effects will be higher. We anticipate that studies focused on these populations are likely to identify additional mutations that affect germline-mutation rate.

We found that, among 7,700 100kGP families, parental age explained only around 70% of the variance in the numbers of dnSNVs per offspring, which is substantially smaller than a previous estimate of 95% based on 78 families¹. Resampling analyses showed that, in small numbers of families, estimates of the variance explained by parental age have wide confidence intervals such that these two estimates are not inconsistent, although estimates based on a two order of magnitude greater number of samples will be much more precise. A residual ~20% of variation in the numbers of germline dnSNVs per individual remains unexplained by parental age, data quality and hypermutation. We found that neither rare variants in known DNA-repair genes nor polygenic contributions from common variants ($MAF > 0.01$) are likely to account for a large proportion of this unexplained variance. Larger sample sizes are required to further evaluate polygenic contributions from intermediate frequency ($0.001 < MAF < 0.01$) variants. A limitation of these heritability analyses is the use of DNMs in offspring as a proxy for germline-mutation rates in individual parents. Measuring germline-mutation rates more directly by, for example, sequencing hundreds of single gametes per individual, should facilitate better powered association studies and heritability analyses.

Environmental exposures are also likely to contribute to germline-mutation rate variation. We have observed evidence that certain chemotherapeutic agents can affect the germline-mutation rate. Targeted studies on the germline mutagenic effects of different chemotherapeutic agents (such as in cancer survivor cohorts) will be crucial in understanding this further. We anticipate heterogeneity in the germline mutagenic effects of different chemotherapeutic agents, in part due to differences in the permeability of the blood–testis barrier⁵⁹, as well as variation in the vulnerability to chemotherapeutic germline mutagenesis by sex and age. As few individuals receive chemotherapy before reproduction, chemotherapeutic exposures will not explain a large proportion of the remaining variation in germline-mutation rates. However chemotherapeutic mutagenesis has important implications for patients receiving some chemotherapies who plan to have

children, especially in relation to storing unexposed gametes for future use of assisted reproductive technologies.

Unexplained hypermutation and additional variance in the germline-mutation rate might be explained by other environmental exposures. One limitation of this study was the lack of data on non-therapeutic environmental exposures. Reassuringly, the narrow distribution of DNMs per individual in the 100kGP cohort suggests that it is unlikely that there are common environmental mutagen exposures in the UK (such as cigarette smoking) that cause a substantive (for example, >1.5 times) fold increase in mutation rates and concomitant disease risk. The germline generally appears to be well protected from large increases in mutation rate. However, including a broader spectrum of environmental exposures in future studies would help to identify more subtle effects and may reveal gene-by-environment interactions.

Methods

DNM filtering in 100,000 Genomes Project

We analysed DNMs called in 13,949 parent–offspring trios from 12,609 families from the rare disease programme of the 100,000 Genomes Project. The rare disease cohort includes individuals with a wide array of diseases, including neurodevelopmental disorders, cardiovascular disorders, renal and urinary tract disorders, ophthalmological disorders, tumour syndromes, ciliopathies and others. These are described in more detail in previous publications^{60,61}. The cohort was whole-genome sequenced at around 35× coverage and variant calling for these families was performed through the Genomics England rare disease analysis pipeline. The details of sequencing and variant calling have been previously described⁶¹. DNMs were called by the Genomics England Bioinformatics team using the Platypus variant caller⁶². These were selected to optimize various properties, including the number of DNMs per person being approximately what we would expect, the distribution of the VAF of the DNMs to be centred around 0.5 and the true positive rate of DNMs to be sufficiently high as calculated from examining IGV plots. The filters applied were as follows:

- Genotype is heterozygous in child (1/0) and homozygous in both parents (0/0).
- Child read depth (RD) > 20, mother RD > 20, father RD > 20.
- Remove variants with >1 alternative read in either parent.
- VAF > 0.3 and VAF < 0.7 for child.
- Remove SNVs within 20 bp of each other. Although this is probably removing true MNVs, the error mode was very high for clustered mutations.
- Removed DNM if child RD > 98 (ref. [14](#)).
- Removed DNM that fell within known segmental duplication regions as defined by the UCSC (<http://humanparalogy.gs.washington.edu/build37/data/GRCh37GenomicSuperDup.tab>).
- Removed DNM that fell in highly repetitive regions (<http://humanparalogy.gs.washington.edu/build37/data/GRCh37simpleRepeat.txt>).
- For DNM calls that fell on the X chromosome, these slightly modified filters were used:
- For DNM that fell in PAR regions, the filters were unchanged from the autosomal calls apart from allowing for both heterozygous (1/0) and hemizygous (1) calls in males.
- For DNM that fell in non-PAR regions the following filters were used:
- For males: RD > 20 in child, RD > 20 in mother, no RD filter on father.
- For males: the genotype must be hemizygous (1) in child and homozygous in mother (0/0).
- For females: RD > 20 in child, RD > 20 in mother, RD > 10 in father.

DNM filtering in DDD

To identify individuals with hypermutation in the DDD study, we started with exome-sequencing data from the DDD study of families with a child with a severe, undiagnosed developmental disorder. The recruitment of these families has been described previously⁶³: families were recruited at 24 clinical genetics centres within the UK National Health Service and the Republic of Ireland. Families gave informed consent to participate, and the study was approved by the UK Research Ethics Committee (10/H0305/83, granted by the Cambridge South Research Ethics Committee, and GEN/284/12, granted by the Republic of Ireland Research Ethics Committee). Sequence alignment and variant calling of SNVs and indels were conducted as previously described. DNMs were called using DeNovoGear and filtered as described previously^{12,64}. The analysis in this paper was conducted on a subset (7,930 parent–offspring trios) of the full current cohort, which was not available at the start of this research.

In the DDD study, we identified 9 individuals out of 7,930 parent–offspring trios with an increased number of exome DNMs after accounting for parental age (7–17 exome DNMs compared to an expected number of ~2). These were subsequently submitted along with their parents for PCR-free whole-genome sequencing at >30x mean coverage using Illumina 150bp paired end reads and in house WSI sequencing pipelines. Reads were mapped with bwa (v0.7.15)⁶⁵. DNMs were called from these trios using DeNovoGear⁶⁴ and were filtered as follows:

- Child RD > 10, mother RD > 10, father RD > 10.
- Alternative allele RD in child of >2.
- Filtered on strand bias across parents and child (p-value > 0.001, Fisher’s exact test).
- Removed DNMs that fell within known segmental duplication regions as defined by the UCSC (<http://humanparalogy.gs.washington.edu/build37/data/GRCh37GenomicSuperDup.tab>).

- Removed DNM s that fell in highly repetitive regions (<http://humanparalogy.gs.washington.edu/build37/data/GRCh37simpleRepeat.txt>).
- Allele frequency in gnomAD < 0.01.
- VAF < 0.1 for both parents.
- Removed mutations if both parents have >1 read supporting the alternative allele.
- Test to see whether VAF in the child is significantly greater than the error rate at that site as defined by error sites estimated using Shearwater⁶⁶.
- Posterior probability from DeNovoGear > 0.00781 (refs. [12,64](#)).
- Removed DNM s if the child RD > 200.

After applying these filters, this resulted in 1,367 DNM s. All of these DNM s were inspected in the Integrative Genome Viewer⁶⁷ and removed if they appeared to be false-positives. This resulted in a final set of 916 DNM s across the 9 trios. One out of the nine had 277 dnSNVs genome wide, whereas the others had expected numbers (median, 81 dnSNVs).

Parental phasing of DNM s

To phase the DNM s in both 100kGP and DDD, we used a custom script that used the following read-based approach to phase a DNM. This first searches for heterozygous variants within 500 bp of the DNM that was able to be phased to a parent (so not heterozygous in both parents and offspring). We next examined the reads or read pairs that included both the variant and the DNM and counted how many times we observed the DNM on the same haplotype of each parent. If the DNM appeared exclusively on the same haplotype as a single parent then that was determined to originate from that parent. We discarded DNM s that had conflicting evidence from both parents. This code is available on GitHub (<https://github.com/queenjobo/PhaseMyDeNovo>).

Parental age and germline-mutation rate

To assess the effect of parental age on germline-mutation rate, we ran the following regressions on autosomal DNMs. These and subsequent statistical analyses were performed primarily in R (v.4.0.1). On all (unphased) DNMs, we ran two separate regressions for SNVs and indels. We chose a negative binomial generalized linear model (GLM) here as the Poisson was found to be overdispersed. We fitted the following model using a negative Binomial GLM with an identity link where Y is the number of DNMs for an individual:

$$E(Y) = \beta_0 + \beta_1 \text{paternal age} + \beta_2 \text{maternal age}$$

For the phased DNMs we fit the following two models using a negative binomial GLM with an identity link where Y_{maternal} is the number of maternally derived DNMs and Y_{paternal} is the number of paternally derived DNMs:

$$E(Y_{\text{paternal}}) = \beta_0 + \beta_1 \text{paternal age}$$

$$E(Y_{\text{maternal}}) = \beta_0 + \beta_1 \text{maternal age}$$

Individuals with hypermutation in the 100kGP cohort

To identify individuals with hypermutation in the 100kGP cohort, we first wanted to regress out the effect of parental age as described in the parental age analysis. We then looked at the distribution of the studentized residuals and then, assuming these followed a t distribution with $N - 3$ degrees of freedom, calculated a t -test P value for each individual. We took the same approach for the number of indels except, in this case, Y would be the number of de novo indels.

We identified 21 individuals out of 12,471 parent–offspring trios with a significantly increased number of dnSNVs genome wide ($P < 0.05/12,471$ tests). We performed multiple quality control analyses, which included examining the mutations in the Integrative Genomics Browser for these individuals to examine DNM calling accuracy, looking at the relative

position of the DNMs across the genome and examining the mutational spectra of the DNMs to identify any well-known sequencing error mutation types. We identified 12 that were not truly hypermutated. The majority of false-positives (10) were due to a parental somatic deletion in the blood, increasing the number of apparent DNMs (Supplementary Fig. 7). These individuals had some of the highest numbers of DNMs called (up to 1,379 DNMs per individual). For each of these 10 individuals, the DNM calls all clustered to a specific region in a single chromosome. In this same corresponding region in the parent, we observed a loss of heterozygosity when calculating the heterozygous/homozygous ratio. Moreover, many of these calls appeared to be low-level mosaic in that same parent. This type of event has previously been shown to create artifacts in CNV calls and is referred to as a ‘loss of transmitted allele’ event⁶⁸. The remaining two false-positives were due to bad data quality in either the offspring or one of the parents leading to poor DNM calls. The large number of DNMs in these false-positive individuals also led to significant underdispersion in the model so, after removing these 12 individuals, we reran the regression model and subsequently identified 11 individuals who appeared to have true hypermutation ($P < 0.05/12,459$ tests).

Extraction of mutational signatures

Mutational signatures were extracted from maternally and paternally phased autosomal DNMs, 24 controls (randomly selected), 25 individuals (father with a cancer diagnosis before conception), 27 individuals (mother with a cancer diagnosis before conception) and 12 individuals with hypermutation that we identified. All DNMs were lifted over to GRCh37 before signature extraction (100kGP samples are a mix of GRCh37 and GRCh38) and, through the liftover process, a small number of 100kGP DNMs were lost (0.09% overall, 2 DNMs were lost across all of the individuals with hypermutation). The mutation counts for all of the samples are shown in Supplementary Table 1. This was performed using SigProfiler (v.1.0.17) and these signatures were extracted and subsequently mapped on to COSMIC mutational signatures (COSMIC v.91, Mutational Signature v.3.1)^{19,40}. SigProfiler defaults to selecting a solution with higher specificity than sensitivity. A solution with 4 de novo signatures was chosen as optimal by SigProfiler for the 12 individuals with germline-hypermutated genomes.

Another stable solution with five de novo signatures was also manually deconvoluted, which has been considered as the final solution. The mutation probability for mutational signature SBSHYP is shown in Supplementary Table 3.

External exposure signature comparison

We compared the extracted signatures from these individuals with hypermutation with a compilation of previously identified signatures caused by environmental mutagens from the literature. The environmental signatures were compiled from refs. [24,51,52](#). Comparison was calculated as the cosine similarity between the different signatures.

Genes involved in DNA repair

We compiled a list of DNA-repair genes that were taken from an updated version of the table in ref. [69](#) (<https://www.mdanderson.org/documents/Labs/Wood-Laboratory/human-dna-repair-genes.html>). These can be found in Supplementary Table 4. These are annotated with the pathways that they are involved with (such as nucleotide-excision repair, mismatch repair). A ‘rare’ variant is defined as those with an allele frequency of <0.001 for heterozygous variants and those with an allele frequency of <0.01 for homozygous variants in both the 1000 Genomes as well as across the 100kGP cohort.

Kinetic characterization of MPG

The A135T variant of *MPG* was generated by site-directed mutagenesis and confirmed by sequencing both strands. The catalytic domain of WT and A135T *MPG* was expressed in BL21(DE3) Rosetta2 *Escherichia coli* and purified as described for the full-length protein [70](#). Protein concentration was determined by absorbance at 280 nm. Active concentration was determined by electrophoretic mobility shift assay with 5'-FAM-labelled pyridine-DNA [48](#) (Extended Data Fig. 8). Glycosylase assays were performed with 50 mM NaMOPS, pH 7.3, 172 mM potassium acetate, 1 mM DTT, 1 mM EDTA, 0.1 mg ml⁻¹ BSA at 37 °C. For single-turnover glycosylase activity, a 5'-FAM-labelled duplex was annealed by heating to 95 °C and slowly

cooling to 4 °C (Extended Data Fig. 9). DNA substrate concentration was varied between 10 nM and 50 nM, and MPG concentration was maintained in at least twofold excess over DNA from 25 nM to 10,000 nM. Samples taken at timepoints were quenched in 0.2 M NaOH, heated to 70 °C for 12.5 min, then mixed with formamide/EDTA loading buffer and analysed by 15% denaturing polyacrylamide gel electrophoresis. Fluorescence was quantified using the Typhoon 5 imager and ImageQuant software (GE). The fraction of product was fit by a single exponential equation to determine the observed single-turnover rate constant (k_{obs}). For Hx excision, the concentration dependence was fit by the equation $k_{\text{obs}} = k_{\text{max}} [\text{E}] / (K_{1/2} + [\text{E}])$, where $K_{1/2}$ is the concentration at which half the maximal rate constant (k_{max}) was obtained and $[\text{E}]$ is the concentration of enzyme. It was not possible to measure the $K_{1/2}$ for εA excision using a fluorescence-based assay owing to extremely tight binding⁷¹. Multiple turnover glycosylase assays were performed with 5 nM MPG and 10–40-fold excess of substrate (Extended Data Fig. 8).

Fraction of variance explained

To estimate the fraction of germline mutation variance explained by several factors, we fit the following negative binomial GLMs with an identity link. Data quality is likely to correlate with the number of DNM s detected so, to reduce this variation, we used a subset of the 100kGP dataset that had been filtered on some base quality control metrics by the Bioinformatics team at GEL:

- Cross-contamination < 5%
- Mapping rate > 75%
- Mean sample coverage > 20
- Insert size < 250

We then included the following variables to try to capture as much of the residual measurement error which may also be impacting DNM calling. In brackets are the corresponding variable names used in the models below:

- Mean coverage for the child, mother and father (child mean RD, mother mean RD, father mean RD)
- Proportion of aligned reads for the child, mother and father (child prop aligned, mother prop aligned, father prop aligned)
- Number of SNVs called for child, mother and father (child snvs, mother snvs, father snvs)
- Median VAF of DNMs called in child (median VAF)
- Median ‘Bayes Factor’ as outputted by Platypus for DNMs called in the child. This is a metric of DNM quality (median BF).

The first model only included parental age:

$$E(Y) = \beta_0 + \beta_1 \text{paternal age} + \beta_2 \text{maternal age}$$

The second model also included data quality variables as described above:

```
 $$\begin{array}{cc} E(Y) = & \beta_0 + \beta_1 \text{paternal age} \\ & + \beta_2 \text{maternal age} \\ & + \beta_3 \text{child mean RD} \\ & + \beta_4 \text{mother mean RD} \\ & + \beta_5 \text{father mean RD} \\ & + \beta_6 \text{child prop aligned} \\ & + \beta_7 \text{mother prop aligned} \\ & + \beta_8 \text{father prop aligned} \\ & + \beta_9 \text{child snvs} \\ & + \beta_{10} \text{mother snvs} \\ & + \beta_{11} \text{father snvs} \\ & + \beta_{12} \text{median VAF} \\ & + \beta_{13} \text{median BF} \end{array} $$
```

The third model included a variable for excess mutations in the 11 confirmed individuals with hypermutation (hm excess) in the 100kGP dataset. This variable was the total number of mutations subtracted by the median number of DNMs in the cohort (65), $Y_{\text{hypermutated}} - \text{median}(Y)$ for these 11 individuals and 0 for all other individuals.

```
 $$\begin{array}{cc} E(Y) = & \beta_0 + \beta_1 \text{paternal age} \\ & + \beta_2 \text{maternal age} \\ & + \beta_3 \text{child mean RD} \\ & + \beta_4 \text{mother mean RD} \\ & + \beta_5 \text{father mean RD} \\ & + \beta_6 \text{child prop aligned} \\ & + \beta_7 \text{mother prop aligned} \\ & + \beta_8 \text{father prop aligned} \\ & + \beta_9 \text{child snvs} \\ & + \beta_{10} \text{mother snvs} \\ & + \beta_{11} \text{father snvs} \\ & + \beta_{12} \text{median VAF} \\ & + \beta_{13} \text{median BF} \\ & + \beta_{14} \text{hm excess} \end{array} $$
```

```

mean\; RD} }+{ \beta } _{ 4 } { \rm{mother}\; mean\; RD} } \\ & +{ \beta } _{ 5 } \backslash , \\
{ \rm{father}\; mean\; RD} }+{ \beta } _{ 6 } { \rm{child}\; prop\; aligned} } \\ & +{ \beta } _{ 7 } { \rm{mother}\; prop\; aligned} }+{ \beta } _{ 8 } { \rm{father}\; prop\; aligned} } \\ & +{ \beta } _{ 9 } { \rm{child}\; snvs} }+{ \beta } _{ 10 } { \rm{mother}\; snvs} }+{ \beta } _{ 11 } { \rm{father}\; snvs} } \\ & & +{ \beta } _{ 12 } { \rm{median}\; VAF} }+{ \beta } _{ 13 } { \rm{median}\; BF} }+{ \beta } _{ 14 } { \rm{hm}\; excess} } \end{array} $$

```

The fraction of variance (F) explained after accounting for Poisson variance in the mutation rate was calculated in a similar way to in ref. [1](#) using the following formula:

$$F = \frac{\text{Var}(\text{Y})}{\text{Var}(\text{pseudo})}$$

McFadden's pseudo R^2 was used here as a negative binomial GLM was fitted. We repeated these analyses fitting an ordinary least squares regression, as was done in ref. [1](#), using the R^2 and got comparable results. To calculate a 95% confidence interval, we used a bootstrapping approach. We sampled with a replacement 1,000 times and extracted the 2.5% and 97.5% percentiles.

Rare variants in DNA-repair genes

We fit eight separate regressions to assess the contribution of rare variants in DNA-repair genes (compiled as described previously). These were across three different sets of genes: variants in all DNA-repair genes, variants in a subset of DNA-repair genes that are known to be associated with base-excision repair, MMR, NER or a DNA polymerase, and variants within this subset that have also been associated with a cancer phenotype. For this, we downloaded all ClinVar entries as of October 2019 and searched for germline 'pathogenic' or 'likely pathogenic' variants annotated with cancer^{[55](#)}. We tested both all non-synonymous variants and just PTVs for each set. To assess the contribution of each of these sets, we created two binary variables per set indicating a presence or absence of a maternal or paternal variant for each individual, and then ran a negative binomial regression for each subset including these as independent variables along

with hypermutation status, parental age and quality-control metrics as described in the previous section.

Simulations for parental age effect

We downsampled from the full cohort to examine how the estimates of the fraction of variance in the number of DNM explained by paternal age varied with sample number. We first simulated a random sample as follows 10,000 times:

- Randomly sample 78 trios (the number of trios in ref. [1](#).)
- Fit ordinary least squares of $E(Y) = \beta_0 + \beta_1$ paternal age.
- Estimated the fraction of variance (F) as described in ref. [1](#).

We found that the median fraction explained was 0.77, with a s.d. of 0.13 and with 95% of simulations fallings between 0.51 and 1.00.

Parental cancer diagnosis before conception

To identify parents who had received a cancer diagnosis before the conception of their child, we examined the admitted patient care hospital episode statistics of these parents. There were no hospital episode statistics available before 1997, and many individuals did not have any records until after the birth of the child. To ensure that comparisons were not biased by this, we first subset to parents who had at least one episode statistic recorded at least two years before the child's year of birth. Two years before the child's birth was our best approximation for before conception without the exact child date of birth. This resulted in 2,891 fathers and 5,508 mothers. From this set we then extracted all entries with ICD10 codes with a 'C' prefix, which corresponds to malignant neoplasms, and 'Z85', which corresponds to a personal history of malignant neoplasm. We defined a parent as having a cancer diagnosis before conception if they had any of these codes recorded ≥ 2 years before the child's year of birth. We also extracted all entries with ICD10 code 'Z511', which codes for an 'encounter for antineoplastic chemotherapy and immunotherapy'.

Two fathers of individuals with hypermutation who we suspect had chemotherapy before conception did not meet these criteria as the father of GEL_5 received chemotherapy for treatment for systemic lupus erythematosus and not cancer and, for the father of GEL_8, the hospital record ‘personal history of malignant neoplasm’ was entered after the conception of the child (Supplementary Table 5).

To compare the number of dnSNVs between the group of individuals with parents with and without cancer diagnoses, we used a Wilcoxon test on the residuals from the negative binomial regression on dnSNVs correcting for parental age, hypermutation status and data quality. To look at the effect of maternal cancer on dnSNVs, we matched these individuals on maternal and paternal age with sampling replacement with 20 controls for each of the 27 individuals. We found a significant increase in DNMs (74 compared to 65 median dnSNVs, $P = 0.001$, Wilcoxon Test).

SNP heritability analysis

For this analysis, we started with the same subset of the 100kGP dataset that had been filtered as described in the analysis of the impact of rare variants in DNA-repair genes across the cohort (see above). To ensure variant quality, we subsetted to variants that have been observed in genomes from gnomAD (v.3)⁷². These were then filtered by ancestry to parent–offspring trios where both the parents and child mapped on to the 1000 Genomes GBR subpopulations. The first 10 principal components were subsequently included in the heritability analyses. To remove cryptic relatedness, we removed individuals with an estimated relatedness of >0.025 (using GCTA grm-cutoff, 0.025). This resulted in a set of 6,352 fathers and 6,329 mothers. The phenotype in this analysis was defined as the residual from the negative binomial regression of the number of DNMs after accounting for parental age, hypermutation status and several data quality variables, as described when estimating the fraction of DNM count variation explained (see above). To estimate heritability, we ran GCTA GREML-LDMS on two linkage disequilibrium stratifications and three MAF bins (0.001–0.01, 0.01–0.05, 0.05–1)⁵⁶. For mothers, this was run with the --reml-no-constrain option because it would otherwise not converge (Supplementary Table 9).

Reporting summary

Further information on research design is available in the [Nature Research Reporting Summary](#) linked to this paper.

Data availability

Sequence and variant-level data and phenotypic data for the DDD study data are available from the European Genome–Phenome Archive (EGA: [EGAS00001000775](#)). The DDD_1 WGS and DNM data are under [EGAD00001008497](#). These data are under managed access to ensure that the work proposed by the researchers is allowed under the study's ethical approval. Sequence- and variant-level data (including the DNM dataset) and phenotypic data from the 100,000 Genomes Project can be accessed by application to Genomics England following the procedure outlined at <https://www.genomicsengland.co.uk/about-gecip/joining-research-community/>. Other databases are available online: Genome Aggregation Database (gnomAD v.2.1.1; <https://gnomad.broadinstitute.org/>); Catalogue of Somatic Mutations in Cancer (v.3.1; <https://cancer.sanger.ac.uk/>); ClinVar (<https://www.ncbi.nlm.nih.gov/clinvar/>).

Code availability

Phasing of mutations was performed with a custom Python (3) script available at GitHub (<https://github.com/queenjobo/PhaseMyDeNovo>).

References

1. Kong, A. et al. Rate of de novo mutations and the importance of father's age to disease risk. *Nature* **488**, 471–475 (2012).
2. Jónsson, H. et al. Parental influence on human germline de novo mutations in 1,548 trios from Iceland. *Nature* **549**, 519–522 (2017).
3. Conrad, D. F. et al. Variation in genome-wide mutation rates within and between human families. *Nat. Genet.* **43**, 712–714 (2011).

4. Roach, J. C. et al. Analysis of genetic inheritance in a family quartet by whole-genome sequencing. *Science* **328**, 636–639 (2010).
5. Sasani, T. A. et al. Large, three-generation human families reveal postzygotic mosaicism and variability in germline mutation accumulation. *eLife* **8**, e46922 (2019).
6. Lindsay, S. J., Rahbari, R., Kaplanis, J., Keane, T. & Hurles, M. E. Similarities and differences in patterns of germline mutation between mice and humans. *Nat. Commun.* **10**, 4053 (2019).
7. Harris, K. Evidence for recent, population-specific evolution of the human mutation rate. *Proc. Natl Acad. Sci. USA* **112**, 3439–3444 (2015).
8. Yang, S. et al. Parent–progeny sequencing indicates higher mutation rates in heterozygotes. *Nature* **523**, 463–467 (2015).
9. Amos, W. Flanking heterozygosity influences the relative probability of different base substitutions in humans. *R. Soc. Open Sci.* **6**, 191018 (2019).
10. Girard, S. L. et al. Paternal age explains a major portion of de novo germline mutation rate variability in healthy individuals. *PLoS ONE* **11**, e0164212 (2016).
11. Mitra, I. et al. Patterns of de novo tandem repeat mutations and their role in autism. *Nature* **589**, 246–250 (2021).
12. Deciphering Developmental Disorders Study. Prevalence and architecture of de novo mutations in developmental disorders. *Nature* **542**, 433–438 (2017).
13. Goldmann, J. M. et al. Parent-of-origin-specific signatures of de novo mutations. *Nat. Genet.* **48**, 935–939 (2016).
14. Rahbari, R. et al. Timing, rates and spectra of human germline mutation. *Nat. Genet.* **48**, 126–133 (2016).

15. Gao, Z. et al. Overlooked roles of DNA damage and maternal age in generating human germline mutations. *Proc. Natl Acad. Sci. USA* **116**, 9491–9500 (2019).
16. Alexandrov, L. B., Nik-Zainal, S., Wedge, D. C., Campbell, P. J. & Stratton, M. R. Deciphering signatures of mutational processes operative in human cancer. *Cell Rep.* **3**, 246–259 (2013).
17. Nik-Zainal, S. et al. 604 cancer genomics, epigenetics and genomic instability. Mutational processes shaping the genomes of twenty-one breast cancers. *Eur. J. Cancer* **48**, S144 (2012).
18. Phillips, D. H. Mutational spectra and mutational signatures: insights into cancer aetiology and mechanisms of DNA damage and repair. *DNA Repair* **71**, 6–11 (2018).
19. Alexandrov, L. B. et al. The repertoire of mutational signatures in human cancer. *Nature* **578**, 94–101 (2020).
20. Demanelis, K. et al. Determinants of telomere length across human tissues. *Science* **369**, eaaz6876 (2020).
21. Moore, L. et al. The mutational landscape of human somatic and germline cells. *Nature* **597**, 381–386 (2021).
22. Alexandrov, L. B. et al. Clock-like mutational processes in human somatic cells. *Nat. Genet.* **47**, 1402–1407 (2015).
23. Stratton, M. R., Campbell, P. J. & Andrew Futreal, P. The cancer genome. *Nature* **458**, 719–724 (2009).
24. Kucab, J. E. et al. A compendium of mutational signatures of environmental agents. *Cell* **177**, 821–836 (2019).
25. Tawn, E. J. & Janet Tawn, E. Hereditary effects of radiation: UNSCEAR 2001 report to the General Assembly, with scientific annex. *J. Radiol. Prot.* **22**, 121–122 (2002).

26. Forster, L., Forster, P., Lutz-Bonengel, S., Willkomm, H. & Brinkmann, B. Natural radioactivity and human mitochondrial DNA mutations. *Proc. Natl Acad. Sci. USA* **99**, 13950–13954 (2002).
27. Adewoye, A. B., Lindsay, S. J., Dubrova, Y. E. & Hurles, M. E. The genome-wide effects of ionizing radiation on mutation induction in the mammalian germline. *Nat. Commun.* **6**, 6684 (2015).
28. Huang, K.-L. et al. Pathogenic germline variants in 10,389 adult cancers. *Cell* **173**, 355–370 (2018).
29. Dutil, J. et al. Germline variants in cancer genes in high-risk non-BRCA patients from Puerto Rico. *Sci. Rep.* **9**, 17769 (2019).
30. Hu, C. et al. Association between inherited germline mutations in cancer predisposition genes and risk of pancreatic cancer. *JAMA* **319**, 2401–2409 (2018).
31. Huang, Q.-Y. et al. Mutation patterns at dinucleotide microsatellite loci in humans. *Am. J. Hum. Genet.* **70**, 625–634 (2002).
32. Gymrek, M., Willem, T., Reich, D. & Erlich, Y. Interpreting short tandem repeat variations in humans using mutational constraint. *Nat. Genet.* **49**, 1495–1501 (2017).
33. Sun, J. X. et al. A direct characterization of human mutation based on microsatellites. *Nat. Genet.* **44**, 1161–1165 (2012).
34. Monckton, D. G. et al. Minisatellite mutation rate variation associated with a flanking DNA sequence polymorphism. *Nat. Genet.* **8**, 162–170 (1994).
35. Crowley, J. J. et al. Common-variant associations with fragile X syndrome. *Mol. Psychiatry* **24**, 338–344 (2019).
36. Kaplanis, J. et al. Evidence for 28 genetic disorders discovered by combining healthcare and research data. *Nature* **586**, 757–762 (2020).

37. Uchimura, A. et al. Germline mutation rates and the long-term phenotypic effects of mutation accumulation in wild-type laboratory mice and mutator mice. *Genome Res.* **25**, 1125–1134 (2015).
38. Cawthon, R. M. et al. Germline mutation rates in young adults predict longevity and reproductive lifespan. *Sci. Rep.* **10**, 10001 (2020).
39. Liu, P. et al. An organismal CNV mutator phenotype restricted to early human development. *Cell* **168**, 830–842 (2017).
40. Tate, J. G. et al. COSMIC: the catalogue of somatic mutations in cancer. *Nucleic Acids Res.* **47**, D941–D947 (2019).
41. Lehmann, A. R., McGibbon, D. & Stefanini, M. Xeroderma pigmentosum. *Orphanet J. Rare Dis.* **6**, 70 (2011).
42. Pippard, E. C., Hall, A. J., Barker, D. J. & Bridges, B. A. Cancer in homozygotes and heterozygotes of ataxia-telangiectasia and xeroderma pigmentosum in Britain. *Cancer Res.* **48**, 2929–2932 (1988).
43. Jager, M. et al. Deficiency of nucleotide excision repair is associated with mutational signature observed in cancer. *Genome Res.* **29**, 1067–1077 (2019).
44. Miccoli, L. et al. The combined effects of xeroderma pigmentosum C deficiency and mutagens on mutation rates in the mouse germ line. *Cancer Res.* **67**, 4695–4699 (2007).
45. Laskowski, R. A., Stephenson, J. D., Sillitoe, I., Orengo, C. A. & Thornton, J. M. VarSite: disease variants and protein structure. *Protein Sci.* **29**, 111–119 (2020).
46. Rentzsch, P., Witten, D., Cooper, G. M., Shendure, J. & Kircher, M. CADD: predicting the deleteriousness of variants throughout the human genome. *Nucleic Acids Res.* **47**, D886–D894 (2019).
47. Glassner, B. J., Rasmussen, L. J., Najarian, M. T., Posnick, L. M. & Samson, L. D. Generation of a strong mutator phenotype in yeast by

- imbalanced base excision repair. *Proc. Natl Acad. Sci. USA* **95**, 9997–10002 (1998).
48. Eyler, D. E., Burnham, K. A., Wilson, T. E. & O'Brien, P. J. Mechanisms of glycosylase induced genomic instability. *PLoS ONE* **12**, e0174041 (2017).
 49. Connor, E. E., Wilson, J. J. & Wyatt, M. D. Effects of substrate specificity on initiating the base excision repair of *N*-methylpurines by variant human 3-methyladenine DNA glycosylases. *Chem. Res. Toxicol.* **18**, 87–94 (2005).
 50. Boot, A. et al. In-depth characterization of the cisplatin mutational signature in human cell lines and in esophageal and liver tumors. *Genome Res.* **28**, 654–665 (2018).
 51. Pich, O. et al. The mutational footprints of cancer therapies. *Nat. Genet.* **51**, 1732–1740 (2019).
 52. Volkova, N. V. et al. Mutational signatures are jointly shaped by DNA damage and repair. *Nat. Commun.* **11**, 2169 (2020).
 53. Rustad, E. H. et al. Timing the initiation of multiple myeloma. *Nat. Commun.* **11**, 1917 (2020).
 54. Hock, H. & Shimamura, A. ETV6 in hematopoiesis and leukemia predisposition. *Semin. Hematol.* **54**, 98–104 (2017).
 55. Landrum, M. J. et al. ClinVar: public archive of interpretations of clinically relevant variants. *Nucleic Acids Res.* **44**, D862–D868 (2016).
 56. Yang, J. et al. Genetic variance estimation with imputed variants finds negligible missing heritability for human height and body mass index. *Nat. Genet.* **47**, 1114–1120 (2015).
 57. Kilpivaara, O. & Aaltonen, L. A. Diagnostic cancer genome sequencing and the contribution of germline variants. *Science* **339**, 1559–1562 (2013).

58. Wong, E. et al. Mbd4 inactivation increases C→T transition mutations and promotes gastrointestinal tumor formation. *Proc. Natl Acad. Sci. USA* **99**, 14937–14942 (2002).
59. Bart, J. et al. An oncological view on the blood-testis barrier. *Lancet Oncol.* **3**, 357–363 (2002).
60. Wheway, G. & Mitchison, H. M., Genomics England Research Consortium. Opportunities and challenges for molecular understanding of ciliopathies—The 100,000 Genomes Project. *Front. Genet.* **10**, 127 (2019).
61. Turro, E. et al. Whole-genome sequencing of rare disease patients in a national healthcare system. *Nature* **583**, 96–102 (2020).
62. Rimmer, A. et al. Integrating mapping-, assembly- and haplotype-based approaches for calling variants in clinical sequencing applications. *Nat. Genet.* **46**, 912–918 (2014).
63. Wright, C. F. et al. Genetic diagnosis of developmental disorders in the DDD study: a scalable analysis of genome-wide research data. *Lancet* **385**, 1305–1314 (2015).
64. Ramu, A. et al. DeNovoGear: de novo indel and point mutation discovery and phasing. *Nat. Methods* **10**, 985–987 (2013).
65. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
66. Gerstung, M., Papaemmanuil, E. & Campbell, P. J. Subclonal variant calling with multiple samples and prior knowledge. *Bioinformatics* **30**, 1198–1204 (2014).
67. Thorvaldsdottir, H., Robinson, J. T. & Mesirov, J. P. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief. Bioinform.* **14**, 178–192 (2013).
68. Redon, R. et al. Global variation in copy number in the human genome. *Nature* **444**, 444–454 (2006).

69. Lange, S. S., Takata, K.-I. & Wood, R. D. DNA polymerases and cancer. *Nat. Rev. Cancer* **11**, 96–110 (2011).
70. Zhang, Y. & O'Brien, P. J. Repair of alkylation damage in eukaryotic chromatin depends on searching ability of alkyladenine DNA glycosylase. *ACS Chem. Biol.* **10**, 2606–2615 (2015).
71. Wolfe, A. E. & O'Brien, P. J. Kinetic mechanism for the flipping and excision of 1,N(6)-ethenoadenine by human alkyladenine DNA glycosylase. *Biochemistry* **48**, 11357–11369 (2009).
72. Karczewski, K. J. et al. The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature* **581**, 434–443 (2020).

Acknowledgements

We thank the families and their clinicians for their participation and engagement, and our colleagues who assisted in the generation and processing of data; M. Stratton, P. Campbell, E. Mitchell, E. Dunstone, H. Martin, K. Chundru, M. Przeworski and J. Korbel for discussions and advice. This research was made possible through access to the data and findings generated by the 100,000 Genomes Project. The 100,000 Genomes Project is managed by Genomics England Limited (a wholly owned company of the Department of Health and Social Care). The 100,000 Genomes Project is funded by the National Institute for Health Research and NHS England. The Wellcome Trust, Cancer Research UK and the Medical Research Council have also funded research infrastructure. The 100,000 Genomes Project uses data provided by patients and collected by the National Health Service as part of their care and support. The DDD study presents independent research commissioned by the Health Innovation Challenge Fund (grant no. HICF-1009-003). The full acknowledgements can be found online (www.ddduk.org/access.html). This research was funded in part by the Wellcome Trust (grant no. 206194). For the purpose of open access, the authors have applied a CC-BY public copyright licence to any author accepted manuscript version arising from this submission. This work was supported by Health Data Research UK, which is funded by the UK Medical Research Council, Engineering and Physical Sciences Research

Council, Economic and Social Research Council, Department of Health and Social Care (England), Chief Scientist Office of the Scottish Government Health and Social Care Directorates, Health and Social Care Research and Development Division (Welsh Government), Public Health Agency (Northern Ireland), British Heart Foundation and Wellcome.

Author information

Affiliations

1. Wellcome Sanger Institute, Wellcome Genome Campus, Hinxton, UK

Joanna Kaplani, Rakesh Sanghvi, Matthew Neville, Petr Danecek, Tim Coorens, Elena Prigmore, Patrick Short, Giuseppe Gallone, Jeremy McRae, Helen Firth, Raheleh Rahbari & Matthew Hurles

2. Department of Biological Chemistry, University of Michigan, Ann Arbor, MI, USA

Benjamin Ide & Patrick O'Brien

3. East Anglian Medical Genetics Service, Cambridge University Hospitals, Cambridge, UK

Jenny Carmichael & Helen Firth

4. North East Thames Regional Genetics Service, Great Ormond Street Hospital, London, UK

Angela Barnicoat

5. Genomics England, London, UK

Loukas Moutsianas & Chris Odhams

Consortia

Genomics England Research Consortium

- Loukas Moutsianas
- & Chris Odhams

Contributions

J.K. and M.H. conceived the project. J.K., C.O., L.M., P.D., E.P., J.M., G.G. and P.S. contributed to the generation and quality control of data. J.K., R.S., M.N., T.C., B.I. and P.O. performed analyses/experiments and contributed to the generation of figures. J.C., A.B. and H.F. provided clinical data and interpretation. M.H., R.R and P.O. provided experimental and analytical supervision. J.K., R.R. and M.H. wrote the manuscript with input from all of the authors. M.H. supervised the project.

Corresponding author

Correspondence to [Matthew Hurles](#).

Ethics declarations

Competing interests

M.H. is a co-founder of, consultant to and holds shares in Congenica, a genetics diagnostic company. L.M. and C.O. are employees of Genomics England Ltd.

Peer review

Peer review information

Nature thanks Aaron Quinlan 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 Parental age and number of DNMs.](#)

(a) Paternal and maternal age against the number of dnInDels. **(b)** Paternal age against number of paternally phased dnSNVs and maternal age against number of maternally phased dnSNVs. Hypermuted individuals are highlighted in pink (11 individuals in 100kGP) and blue (DDD individuals).

[Extended Data Fig. 2 Mutational spectra and signatures for maternal vs paternal DNMs across 100kGP cohort.](#)

(a) Mutational spectra for maternal vs paternal DNMs across 100kGP cohort (48,381 maternal DNMs and 167,558 paternal DNMs). Significant differences (chi-squared test, two sided, Bonferroni corrected threshold of $P < 0.05/7$) are marked with * (p-values: C > A 4.6310-23, C > G 0.20, C > T 3.2510-80, CpG > TpG 0.75, T > A 0.98, T > C 1.6210-5, T > G 6.8110-28). The 95% confidence intervals are shown. **(b)** Mutational signature decomposition for DNMs in maternally and paternally derived DNMs. Signatures extracted with SigProfiler. Colours correspond to COSMIC signatures.

[Extended Data Fig. 3 Mutational spectra for the DNMs of hypermutated individuals part 1.](#)

(a–f correspond to individual GEL_1, GEL_2, DDD_1, GEL_3, GEL_4, and GEL_5 respectively). Each row is a hypermutated individual showing the mutational spectra according to count of mutations per each single base change (with CpG>TpG mutations separated from other C>T mutations) and the second plot is the mutation count for all 96 mutations in their trinucleotide context. The x-axis demonstrates the reference trinucleotide

sequence with the mutated base highlighted. The colour and label on the bar above indicates the mutation type.

Extended Data Fig. 4 Mutational spectra for the DNMs hypermutated individuals part 2.

(**a–f** correspond to individual GEL_6, GEL_7, GEL_8, GEL_9, GEL_10 and GEL_11 respectively). Each row is a hypermutated individual showing the mutational spectra according to count of mutations per each single base change (with CpG>TpG mutations separated from other C>T mutations) and the second plot is the mutation count for all 96 mutations in their trinucleotide context. The x-axis demonstrates the reference trinucleotide sequence with the mutated base highlighted. The colour and label on the bar above indicates the mutation type.

Extended Data Fig. 5 Novel mutational signature SBSHYP.

Trinucleotide context mutational profile of novel extracted mutational signature SBSHYP.

Extended Data Fig. 6 Transcriptional strand bias for DNMs in hypermutated individuals.

Plot shows the count of each mutation type on the transcribed and untranscribed strand for each individual. P-values of transcriptional strand bias tests are given in Extended Data Table [1](#).

Extended Data Fig. 7 Distribution of VAF for DNMs in hypermutated individuals.

The vertical line indicates 0.5 VAF. The two plots highlighted in pink are those where the DNMs appear post-zygotic. P-values of VAF tests are given in Extended Data Table [1](#).

Extended Data Fig. 8 Determination of active concentration of MPG.

(a) Representative native gel electrophoresis with 20 nM pyrrolidine-DNA (Y•T) and varying concentration of WT or A135T *MPG* (25 mM NaHEPES pH 7.5, 100 mM NaCl, 5% v/v glycerol, 1 mM EDTA, 1 mM DTT). Agarose gels (2% w/v) were run in 0.5X TBE buffer at 10 V/cm at 4 °C. **(b)** Independent dilutions were fit to a binding titration to yield an active fraction of 0.57 for both WT and A135T ($n = 3$). This demonstrates that equal concentrations of WT and A135T were tested in the glycosylase assays. The concentrations listed are not corrected by this factor. The points shown are the mean and error bars show 1 standard deviation.

Extended Data Fig. 9 In vitro glycosylase activity of WT and A135T MPG.

(a) Glycosylase assay for recombinant protein and 25mer lesion-containing oligonucleotides (O'Brien 2003). *MPG* excises lesion X from X•Y duplex to create an abasic site, which is subsequently hydrolysed by NaOH to create a 12mer product. (b) Representative denaturing gel scanned for fluorescein fluorescence. (c-d) Concentration independent excision of εA from opposing T and C shows increased rate of N-glycosidic bond cleavage by A135T. (panel c, $n = 6$; panel d, $n = 4$) (e-f) Concentration dependence for single-turnover excision of Hx from opposing T and C contexts shows decreased catalytic efficiency for A135T as compared to WT *MPG*. These single turnover rate constants were fit to the equation $k_{obs} = k_{max} [MPG] / (K_1/2 + [MPG])$. (g-h) Steady state concentration dependence for excision of εA was performed in order to measure the catalytic efficiency (k_{cat}/K_M) for A135T and WT *MPG* using 5 nM enzyme and the indicated concentration of substrate. To circumvent the tight binding by *MPG*, 800 mM NaCl was added to the standard buffer as previously described, using the equation $V/E = k_{cat}/K_M[S]$ (panel e-h, $n = 3$). Mean ± SD is shown for at least 3 independent experiments.

Extended Data Table 1 Properties and possible hypermutation sources for germline hypermutated individuals

Supplementary information

Supplementary Information

Supplementary Figs. 1–7, Supplementary Tables 2 and 5–9, descriptions of Supplementary Tables 1, 3 and 4, Supplementary References and Supplementary Notes, detailing the GEL consortia members.

Reporting Summary

Peer Review File

Supplementary Table 1

Supplementary Table 2

Supplementary Table 4

Rights and permissions

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

Reprints and Permissions

About this article

Cite this article

Kaplanis, J., Ide, B., Sanghvi, R. *et al.* Genetic and chemotherapeutic influences on germline hypermutation. *Nature* **605**, 503–508 (2022). <https://doi.org/10.1038/s41586-022-04712-2>

- Received: 01 June 2021
- Accepted: 31 March 2022
- Published: 11 May 2022
- Issue Date: 19 May 2022
- DOI: <https://doi.org/10.1038/s41586-022-04712-2>

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/s41586-022-04712-2>

[Download PDF](#)

- Article
- Open Access
- [Published: 08 April 2022](#)

Phage anti-CBASS and anti-Pycsar nucleases subvert bacterial immunity

- [Samuel J. Hobbs^{1,2},](#)
- [Tanita Wein³,](#)
- [Allen Lu^{1,2},](#)
- [Benjamin R. Morehouse](#) ORCID: [orcid.org/0000-0003-3352-5463^{1,2}](https://orcid.org/0000-0003-3352-5463),
- [Julia Schnabel^{1,2},](#)
- [Azita Leavitt³,](#)
- [Erez Yirmiya³,](#)
- [Rotem Sorek](#) ORCID: [orcid.org/0000-0002-3872-4982³](https://orcid.org/0000-0002-3872-4982) &
- [Philip J. Kranzusch](#) ORCID: [orcid.org/0000-0002-4943-733X^{1,2,4}](https://orcid.org/0000-0002-4943-733X)

[Nature](#) volume 605, pages 522–526 (2022)

- 7640 Accesses
- 1 Citations
- 83 Altmetric
- [Metrics details](#)

Subjects

- [Bacteriophages](#)
- [Hydrolases](#)
- [X-ray crystallography](#)

Abstract

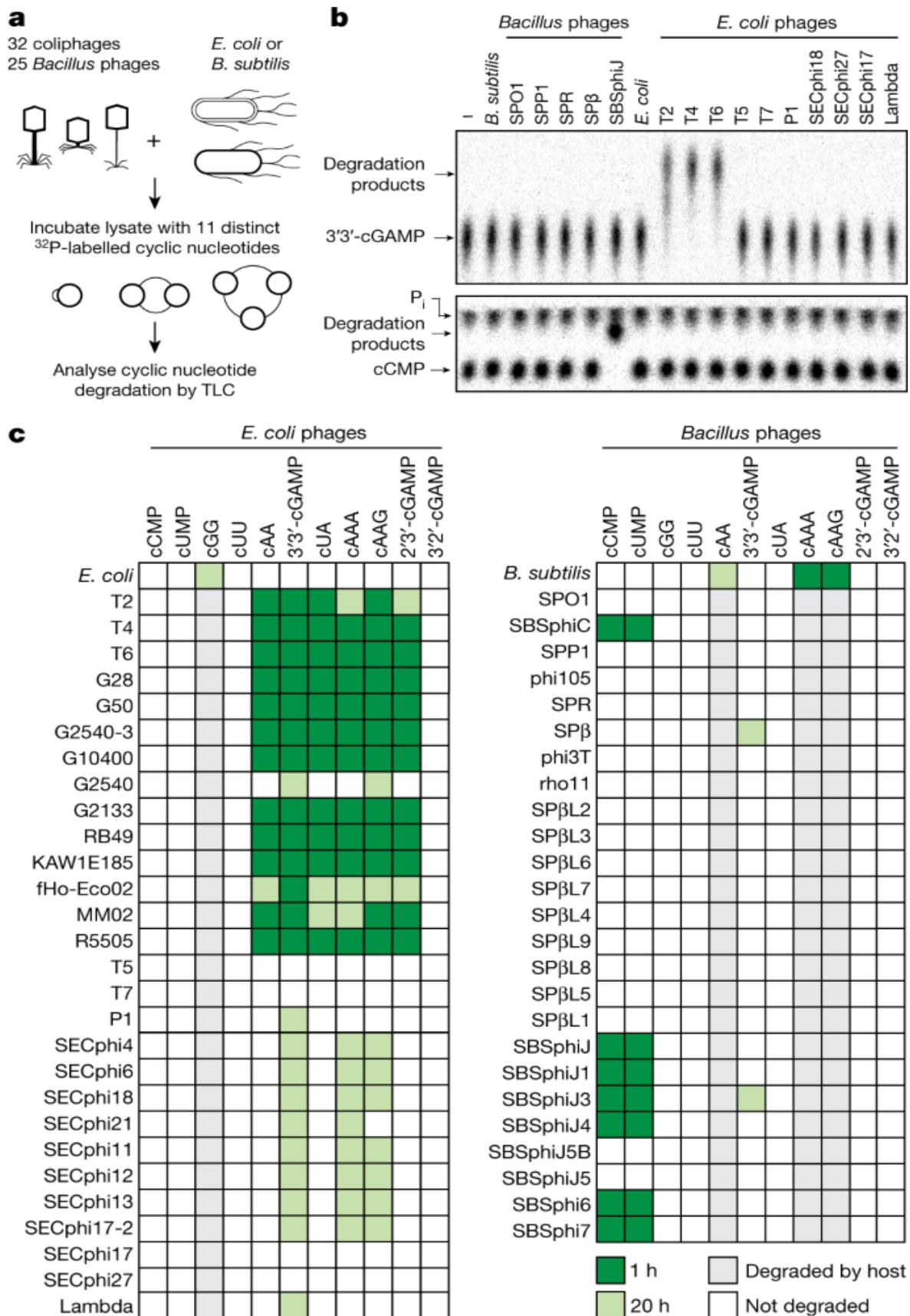
The cyclic oligonucleotide-based antiphage signalling system (CBASS) and the pyrimidine cyclase system for antiphage resistance (Pycsar) are antiphage defence systems in diverse bacteria that use cyclic nucleotide signals to induce cell death and prevent viral propagation^{1,2}. Phages use several strategies to defeat host CRISPR and restriction-modification systems^{3,4,5,6,7,8,9,10}, but no mechanisms are known to evade CBASS and Pycsar immunity. Here we show that phages encode anti-CBASS (Acb) and anti-Pycsar (Apyc) proteins that counteract defence by specifically degrading cyclic nucleotide signals that activate host immunity. Using a biochemical screen of 57 phages in *Escherichia coli* and *Bacillus subtilis*, we discover Acb1 from phage T4 and Apyc1 from phage SBSphiJ as founding members of distinct families of immune evasion proteins. Crystal structures of Acb1 in complex with 3'3'-cyclic GMP–AMP define a mechanism of metal-independent hydrolysis 3' of adenine bases, enabling broad recognition and degradation of cyclic dinucleotide and trinucleotide CBASS signals. Structures of Apyc1 reveal a metal-dependent cyclic NMP phosphodiesterase that uses relaxed specificity to target Pycsar cyclic pyrimidine mononucleotide signals. We show that Acb1 and Apyc1 block downstream effector activation and protect from CBASS and Pycsar defence in vivo. Active Acb1 and Apyc1 enzymes are conserved in phylogenetically diverse phages, demonstrating that cleavage of host cyclic nucleotide signals is a key strategy of immune evasion in phage biology.

Main

To determine how phages evade cyclic nucleotide-based bacterial immune systems, we developed a biochemical screen to analyse the stability of 11 distinct cyclic nucleotide signals during infection with 57 diverse phages (Fig. [1a](#) and Supplementary Table [1](#)). CBASS and Pycsar systems are widely distributed throughout the bacterial kingdom and are present in both Gram-

negative and Gram-positive bacteria including *E. coli* and *B. subtilis*^{1,2,11}. Lysates from uninfected laboratory strains of *E. coli* or *B. subtilis* readily hydrolyse common cyclic nucleotide signals including cyclic-di-GMP (cGG) and cyclic-di-AMP (cAA; Fig. 1b,c and Extended Data Figs. 1 and 2), consistent with known bacterial enzymes that regulate these signals during basal cellular function^{12,13}. By contrast, CBASS and Pycsar antiphage signals including 3'3'-cyclic GMP–AMP (3'3'-cGAMP) and 3',5'-cyclic CMP (cCMP) are exceptionally stable and remain intact following 20-h incubation in uninfected lysates. Strikingly, infection with diverse phages causes rapid hydrolysis of cyclic nucleotide signals specifically involved in immune defence (Fig. 1b,c). Lysates from cells infected with phage T4 and other closely related T-even coliphages degrade distinct classes of CBASS signals including cyclic dinucleotides 3'3'-cGAMP and 3'3'-cyclic UMP–AMP (cUA), and cyclic trinucleotides 3'3'3'-cyclic AMP–AMP–AMP (cAAA) and 3'3'3'-cyclic AMP–AMP–GMP (cAAG). Likewise, lysates from cells infected with the SBSphiJ family of *B. subtilis* phages rapidly degrade the Pycsar signals cCMP and cUMP (Fig. 1b,c). Except for the rare CBASS dinucleotides 3'3'-c-di-UMP and 3'2'-cGAMP, all known cyclic nucleotide signals used in CBASS or Pycsar immune defence were susceptible to degradation by at least one phage (Fig. 1c).

Fig. 1: Phages selectively degrade cyclic nucleotide signals used in host defence.



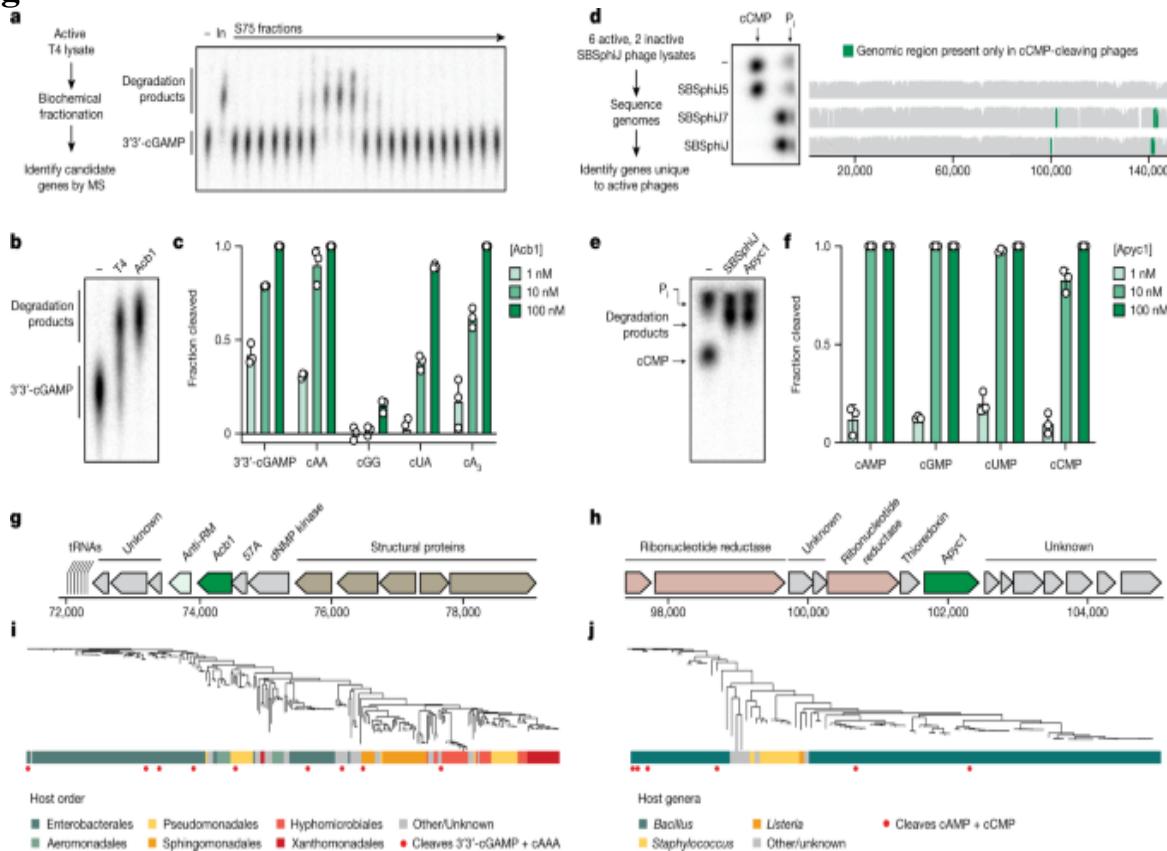
a, Schematic depicting a screen of cyclic nucleotide degradation activity in phage-infected lysates using thin-layer chromatography (TLC). **b**, Representative TLC assays depicting cleavage of 3'3'-cGAMP following infection by T2, T4 and T6 phages, or cleavage of cCMP following infection with SBSphiJ phage. Data are representative of at least two independent replicates. P_i, inorganic phosphate; –, buffer only control. **c**, Summary of the complete results of the screen in **b**, with four phages closely related to T5 omitted for clarity (see Supplementary Table 1 for complete list of phages). The green shading represents the incubation times indicated in the key. T4-related phages degrade diverse CBASS signals and SBSphiJ-related phages degrade diverse Pycsar signals.

Phages encode immune evasion nucleases

Rapid degradation of cyclic nucleotide signals used in host immunity suggests that phages encode proteins dedicated to CBASS and Pycsar evasion. To define anti-CBASS (Acb) and anti-Pycsar (Apyc) proteins, we first focused on phage T4 and used an activity-guided fractionation and mass spectrometry approach to identify candidate Acb proteins responsible for 3'3'-cGAMP cleavage (Fig. 2a and Extended Data Fig. 3a). In vitro screening of each candidate demonstrated that the uncharacterized T4 gene 57B encodes a protein that degrades the CBASS signal 3'3'-cGAMP (Extended Data Fig. 3b,c), and we named this anti-CBASS protein Acb1 (GenBank accession number [NP_049750.1](#)). Recombinant T4 Acb1 rapidly degrades the CBASS signals 3'3'-cGAMP, cUA and cAAA, but does not cleave cGG, demonstrating that Acb1 is responsible for the broad cyclic nucleotide hydrolysis activity observed in T4-infected cell lysate (Fig. 2b,c and Extended Data Fig. 4). We next identified candidate Apyc proteins within SBSphiJ-family phages that cleaved cCMP in our biochemical screen. Genome sequencing and comparative bioinformatic analysis of eight closely related SBSphiJ-family phages revealed two genomic regions present exclusively in phages capable of degrading cCMP (Fig. 2d and Extended Data Fig. 5a). We used structure prediction to analyse each protein encoded in these regions and identified that the uncharacterized SBSphiJ gene 147 encodes a protein with predicted homology to known metallo β-lactamase (MBL) fold RNase and phosphodiesterase enzymes (Fig. 2d and Extended Data Fig. 5a,b). Recombinant protein produced from gene 147

rapidly degrades the Pycsar signals cCMP and cUMP (Fig. 2e,f and Extended Data Fig. 5c–f), and we named this anti-Pycsar protein Apyc1 (European Nucleotide Archive genome accession number [ERS1981056](#)). SBSphiJ Apyc1 efficiently hydrolyses a wide range of cyclic mononucleotides (Fig. 2f), exhibiting an atypically relaxed nucleobase specificity that enables targeting of cyclic pyrimidine signals used in Pycsar immunity.

Fig. 2: Distinct viral nucleases target CBASS and Pycsar immune signals.



a, Schematic and representative example of activity-guided biochemical fractionation and mass spectrometry (MS) to identify Acb1 candidate genes from phage T4. Fractions were collected from an S75 size-exclusion column and tested for 3'3'-cGAMP activity. In, crude lysate input. Data are representative of two independent experiments. **b**, Comparison of 3'3'-cGAMP cleavage by T4 lysate and recombinant Acb1. Data are representative of three independent experiments. **c**, Summary of HPLC analysis testing Acb1 substrate specificity (20-min incubation). Acb1

cleaves dinucleotide and trinucleotide CBASS signals containing one or more AMP. Data are presented as mean \pm s.d. from $n = 3$ independent experiments. **d**, Bioinformatic analysis identifies candidate Apyc1 genes from genomic regions exclusive to cCMP-cleaving phages. TLC data are representative of two independent experiments. **e**, Comparison of cCMP cleavage by SBSphiJ lysate and recombinant Apyc1. Data are representative of three independent experiments. **f**, Summary of HPLC analysis testing Apyc1 substrate specificity (20-min incubation). Apyc1 cleaves all cNMP signals with equal efficiency. Data are presented as mean \pm s.d. from $n = 3$ independent experiments. **g, h**, Schematics showing genes neighbouring T4 Acb1 (**g**) and SBSphiJ Apyc1 (**h**); dNMP, deoxyribosenucleoside monophosphate. **i**, Phylogenetic tree showing T4 Acb1 and 271 related protein sequences from phages, including 112 sequences derived from prophages. Colour strips indicate the order of the bacterial host. Red circles indicate proteins tested for cleavage of 3'3'-cGAMP and cAAA. **j**, Phylogenetic tree displaying SBSphiJ Apyc1 and 106 related protein sequences from phages. Colour strips indicate the genus of the bacterial host. Red circles indicate proteins tested for cleavage of cAMP and cCMP.

Source Data

Immune evasion genes frequently cluster together in the genomes of phages to form anti-defence islands^{7,14}. Consistent with a role in CBASS evasion, T4 Acb1 is encoded adjacent to internal protein I (ipi), a phage inhibitor required to evade the *E. coli* restriction enzyme gmrs/gmrD that recognizes glucosylated cytosine bases present in T4 genomic DNA (ref. 15; Fig. 2g). *Apyc1* is the first identified anti-defence gene in SBSphiJ, limiting comparative analysis with other genes in this phage. However, Apyc1 is encoded adjacent to a series of small proteins of unknown function, suggesting that this variable locus in SBSphiJ-family phages may contribute to evasion of other antiphage defence systems (Fig. 2h). To discover further Acb and Apyc proteins, we searched for proteins related to Acb1 and Apyc1 within phage genomes and prophage sequences (Fig. 2i,j). Analysis of T4 Acb1 identified 281 related protein sequences with about 97% predicted to be of phage origin. We cloned and tested a further 9 *acb1* genes and observed that each recombinant Acb1 protein efficiently cleaved the CBASS signals 3'3'-cGAMP and cAAA (Fig. 2i and Extended Data Fig. 6a). We identified 107 proteins related to Apyc1 present in phage genomes (Fig. 2j)

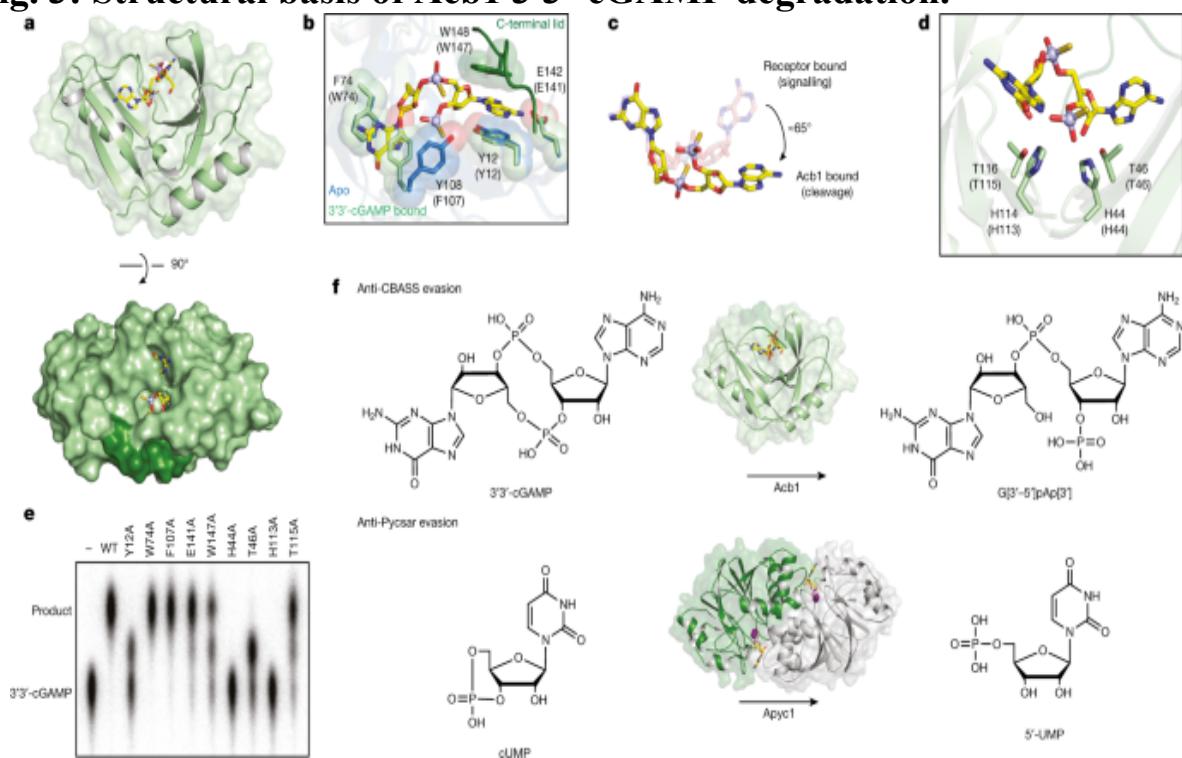
and also found many closely related bacterial proteins encoded in diverse bacterial orders (Extended Data Fig. 6b). Similar to SBSphiJ Apyc1, closely related phage and bacterial Apyc1-like proteins cleaved cyclic mononucleotides with broad specificity (Fig. 2j and Extended Data Fig. 6c). By contrast, the closely related *B. subtilis* enzymes YhfI (GenBank accession number [NP_388905.1](#)) and MBL phosphodiesterase (GenBank accession number [WP_013351727.1](#)) exhibited a strong preference for cAMP/cGMP over cCMP/cUMP cleavage, confirming that relaxed nucleotide specificity and Pycsar signal degradation are unique to Apyc1 and not general features of MBL phosphodiesterase enzymes (Extended Data Fig. 6d). The observation of Apyc1 homologues encoded in bacteria may be explained by the presence of cryptic prophages present in bacterial genomes, but also raises the intriguing possibility that host Apyc1 enzymes may play a role in regulating Pycsar defence or other cNMP-based signalling systems. In total, our analysis identified 273 Acb1 and 107 Apyc1 phage proteins, demonstrating that cyclic nucleotide-degrading enzymes constitute a widespread form of anti-CBASS and anti-Pycsar evasion.

Mechanisms of cyclic nucleotide cleavage

We next determined crystal structures of Acb1 to define the mechanism of anti-CBASS evasion. Structures of Acb1 from the *Erwinia* phage FBB1 in the apo state (1.1 Å) and in complex with 3'3'-cGAMP (1.2 Å) reveal that Acb1 adopts a compact 2H phosphoesterase fold with six central β-strands that form a U-shaped ligand-binding pocket (Fig. 3a, Extended Data Fig. 7a and Supplementary Table 2). On substrate recognition, the flexible carboxy-terminal residues 145–152 form an ordered lid that closes over the top of the captured 3'3'-cGAMP ligand (Fig. 3a and Extended Data Fig. 7b). Acb1 ligand recognition is primarily independent of base identity, with the conserved aromatic residues Y12, W74, F107 and W147 forming stacking interactions with the face of each nucleobase (Fig. 3b). However, base-specific contact occurs between E141 and the 3'3'-cGAMP adenosine N6 position, explaining why at least one adenine is required for cleavage (Fig. 2c and Extended Data Fig. 7c). Although overall lack of sequence-specific contacts allows Acb1 to target a broad range of CBASS cyclic nucleotide signals, the Acb1 binding pocket can accommodate only cyclic dinucleotide or trinucleotide species. Structural clashes prevent recognition of larger

cyclic oligonucleotides with >3 bases, and we confirmed that Acb1 is unable to degrade cyclic tetra-adenylate (cA_4) rings common in type III clustered regularly interspaced short palindromic repeats (CRISPR) immunity^{16,17} (Extended Data Fig. 7d). Acb1–nucleotide interactions contort 3'3'-cGAMP into a highly strained conformation in which the adenine base is rotated about 65° relative to the in-solution or receptor-bound conformation, repositioning the 2' OH for attack on the 3'-5' bond^{18,19} (Fig. 3c). In the Acb1–3'3'-cGAMP structure, the scissile phosphate is positioned over an active-site HxT/HxT tetrad (H44, T46, H113, T115) for acid–base catalysis and the ligand is fully hydrolysed into the linear product G[3'-5']pAp[3'] (GpAp) (Fig. 3d and Extended Data Fig. 7e). We tracked cleavage reactions in vitro using high-performance liquid chromatography (HPLC) and confirmed that Acb1 cleaves 3' of adenosine residues in a two-step, metal-independent reaction that proceeds through a cyclic phosphate intermediate (Extended Data Fig. 7f). Substitutions of conserved active-site and nucleotide-coordinating residues disrupt enzyme function and highlight the critical role for contacts stabilizing the rotated adenine base in Acb1 cyclic nucleotide cleavage (Fig. 3e).

Fig. 3: Structural basis of Acb1 3'3'-cGAMP degradation.



a, Overview of Acb1 from *Erwinia* phage FBB1 in complex with a hydrolysis-resistant phosphorothioate analogue of 3'3'-cGAMP. In the surface representation, C-terminal lid residues are coloured dark green. **b**, Detailed view of residues interacting with the bases of 3'3'-cGAMP. Parentheses indicate equivalent position in T4 Acb1. **c**, Conformations of 3'3'-cGAMP bound to STING (Protein Data Bank (PDB): [5CFM](#)) or Acb1. **d**, Detailed view of catalytic residues. Parentheses indicate equivalent position in T4 Acb1. **e**, Thin-layer chromatography analysis of 3'3'-cGAMP cleavage by T4 Acb1 point mutants. Data are representative of three independent experiments. **f**, Schematic of reactions catalysed by Acb1 and Apyc1.

To compare mechanisms of anti-CBASS and anti-Pycsar evasion, we determined the crystal structure of Apyc1 from the phage Bsp38 (2.7 Å) as well as structures of *Paenibacillus* Apyc1 proteins (1.5 Å and 1.8 Å). These structures confirm that Apyc1 is a member of the class II phosphodiesterase enzymes, which exhibit an MBL fold and have no structural or mechanistic homology to Acb1 (ref. [20](#); Extended Data Fig. [8a,b](#) and Supplementary Table [2](#)). Similar to other structurally characterized class II phosphodiesterases such as *B. subtilis* YhfI, yeast *Saccharomyces cerevisiae* PDE1 or widely distributed RNase Z proteins^{[21,22](#)}, Apyc1 is a homodimer with a highly conserved HxHxDH motif that coordinates two Zn²⁺ ions that bind phosphate groups to position cyclic nucleotides for cleavage (Extended Data Fig. [8a–c](#)). In a structure of *Paenibacillus* Apyc1 co-crystallized in the presence of nonhydrolysable cAMP, we observed strong electron density near the Zn²⁺ ions and more diffuse density in the nucleobase pocket, consistent with specific coordination of the phosphate and ribose backbone of cyclic mononucleotides and weaker nucleobase specificity within the enzyme active site (Extended Data Fig. [8d](#)). Structural comparison of Apyc1 and *B. subtilis* YhfI also reveals that Apyc1 enzymes contain an extended loop that reaches into the nucleotide-binding pocket, potentially enabling stable binding of smaller cyclic pyrimidine substrates (Extended Data Fig. [8b](#)). We confirmed the critical role for Apyc1 metal-coordinating residues and identified E74 and Y112 from the opposing protomer as further catalytic residues required for cCMP hydrolysis and release of the reaction product 5'-CMP (Extended Data Fig. [8e,f](#)). Together, these findings demonstrate that Acb1 and Apyc1 constitute separate families of immune evasion proteins

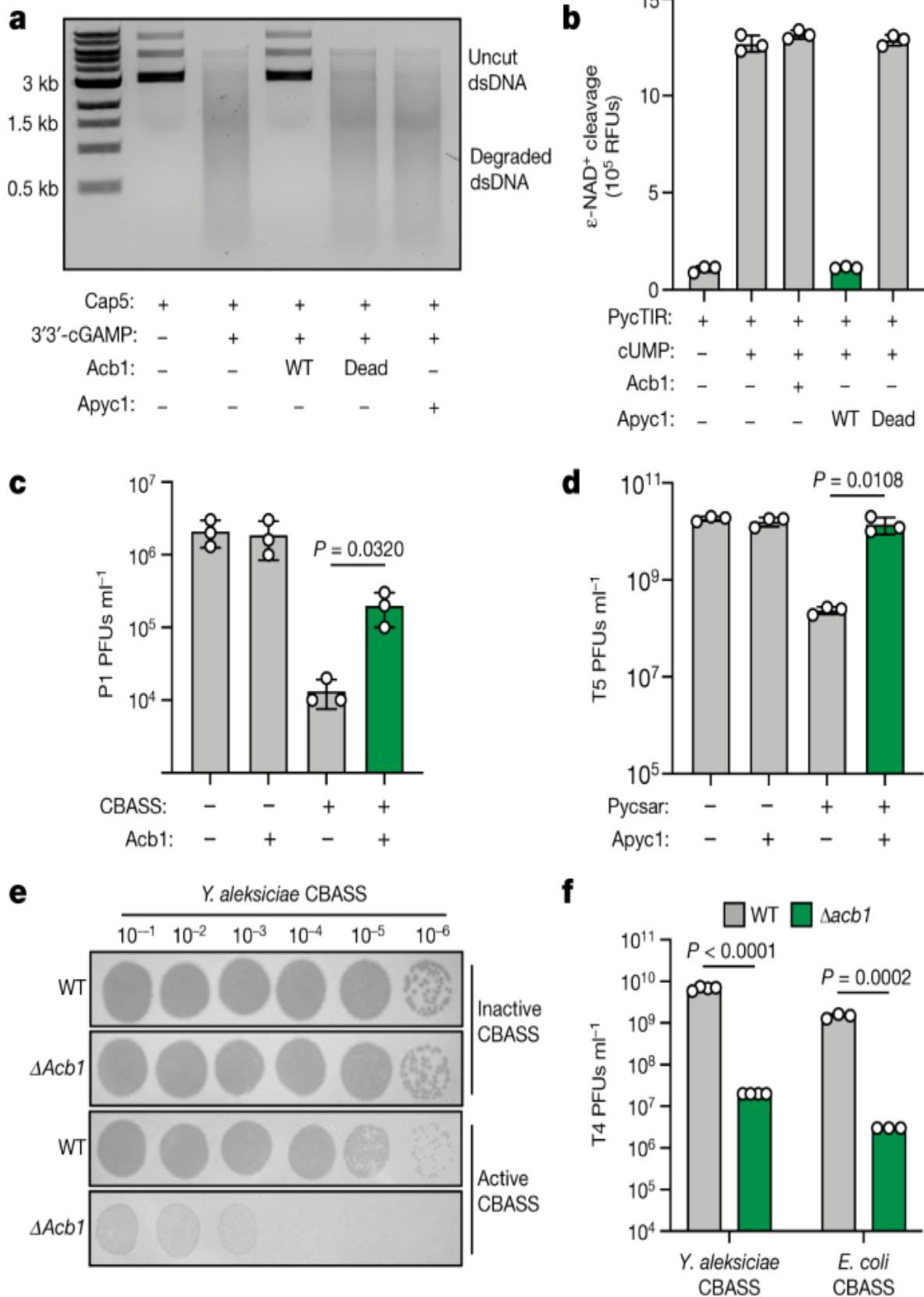
and explain the distinct reaction mechanisms that degrade CBASS or Pycsar cyclic nucleotide signals (Fig. 3f).

Acb1 and Apyc1 subvert host immunity

CBASS and Pycsar antiphage defence requires cyclic nucleotide-dependent activation of downstream effector proteins that induce cell death^{1,2,23,24,25,26}. Using a panel of CBASS nuclease and phospholipase effectors from *Vibrio cholerae*, *Enterobacter cloacae* and *Burkholderia pseudomallei*, we reconstituted CBASS signalling in vitro and observed that Acb1 potently inhibited activation of both cyclic dinucleotide- and cyclic trinucleotide-responsive effectors²³ (Fig. 4a and Extended Data Fig. 9a, b). Likewise, Apyc1 enzymatic activity abolished cUMP-dependent activation of the Pycsar NADase effector PycTIR (ref. 2; Fig. 4b). The activities of Acb1 and Apyc1 are specific to CBASS or Pycsar signalling, demonstrating that anti-CBASS and anti-Pycsar immune evasion proteins are dedicated to each class of antiphage defence system (Fig. 4a, b and Extended Data Fig. 9a).

Tracking Acb1 and Apyc1 activity during infection, we observed that cyclic nucleotide degradation activity begins about 15 min into phage T4 infection and about 30 min into phage SBSphiJ infection, coinciding with the known late onset of CBASS and Pycsar antiphage cell death responses^{1,2} (Extended Data Fig. 9c, d). Acb1 expression in *E. coli* inhibited CBASS-mediated cell death in vivo, suggesting that immune evasion proteins can protect phages from premature abortive infection responses (Extended Data Fig. 9e).

Fig. 4: Acb1 and Apyc1 disrupt CBASS and Pycsar host defence.



a, Agarose gel analysis of uncut plasmid DNA incubated with the CBASS effector Cap5 and 3'3'-cGAMP that was treated with wild-type (WT) Acb1, catalytically inactive Acb1-H44A/H113A or WT Apyc1. Data are representative of three independent experiments. For unprocessed gels, see Supplementary Fig. 1. ds, double-stranded DNA. **b**, Release of fluorescent substrate from an NAD⁺ analogue incubated with the Pycsar effector PycTIR and cUMP that was treated with WT Acb1, WT Apyc1 or catalytically inactive Apyc1-H64A/H66A/H69A. Data are presented as mean ± s.d. from $n = 3$ independent experiments. RFUs, relative fluorescence units. **c**, *E. coli* carrying plasmids encoding a type III CBASS operon from *E. coli* KTE188 and/or T4 Acb1 were challenged with serial dilutions of P1 phage. Data are presented as mean ± s.d. from $n = 3$ independent experiments. PFUs, plaque-forming units. **d**, *E. coli* carrying plasmids encoding a Pycsar operon and/or SBSPhiJ Apyc1 were challenged with serial dilutions of T5 phage. Data are presented as mean ± s.d. from $n = 3$ independent experiments. **e**, Representative plaque assays of *E. coli* carrying a plasmid encoding an active or catalytically inactive CBASS operon from *Yersinia aleksiciae* and challenged with WT phage T4 or phage T4 engineered to remove Acb1 ($\Delta acb1$). **f**, Summary of plaque assay results of WT or $\Delta acb1$ phage T4 infection of *E. coli* carrying CBASS operons from *Y. aleksiciae* or *E. coli*. Data are presented as mean ± s.d. from $n = 4$ (*Y. aleksiciae* operon) or $n = 3$ (*E. coli* operon) technical replicates and are representative of at least 3 biologically independent experiments. Statistical significance in **c**, **d** and **f** was determined using an unpaired two-tailed *t*-test.

[Source Data](#)

To define the importance of degradation of cyclic nucleotide immune signals during phage infection, we infected *E. coli* expressing complete CBASS and Pycsar defence operons and quantified the effect of Acb1 and Apyc1 expression on phage replication. In the presence of an active type III CBASS operon from *E. coli* KTE188, Acb1 expression significantly boosted infectivity of the normally susceptible phage P1 by about 1.5 log (Fig. 4c). Likewise, expression of Apyc1 in *E. coli* disrupted Pycsar defence and completely rescued growth of phage T5, demonstrating that Acb1 and Apyc1 are sufficient to counteract host CBASS and Pycsar defence (Fig. 4d). To determine whether cyclic nucleotide degradation is necessary for immune evasion, we next focused on engineering a mutant phage lacking the

ability to cleave immune nucleotide signals. Robust approaches do not yet exist for genetic manipulation of *B. subtilis* phages, and analysis of *apyc1*-deletion viruses will therefore be a focus of future research. However, we were able to use recent advances in coliphage engineering to create a phage T4 mutant virus lacking functional Acb1 (phage T4 $\Delta acb1$) (Extended Data Fig. 10a). *E. coli* cells infected with phage T4 $\Delta acb1$ do not hydrolyse 3'3'-cGAMP, confirming that Acb1 is essential for viral degradation of CBASS immune cyclic nucleotides (Extended Data Fig. 10b). In the absence of functional CBASS defence, phage T4 and phage T4 $\Delta acb1$ grow equally well, revealing that Acb1 is not required for normal replication in *E. coli* (Fig. 4e,f and Extended Data Fig. 10c,d). In contrast, growth of phage T4 $\Delta acb1$ is specifically impaired in the presence of active CBASS immunity with the mutant virus exhibiting a >300-fold defect in viral replication compared to wild-type phage T4 (Fig. 4e,f and Extended Data Fig. 10c,d). These results demonstrate that viral nucleases are critical for evasion of cyclic nucleotide-mediated phage defence.

Together, our data define Acb1 and Apyc1 as founding members of families of anti-CBASS and anti-Pycsar immune evasion proteins that allow phages to selectively hydrolyse cyclic nucleotide immune signals used for host defence. No single phage could degrade all cyclic nucleotide immune signals, revealing that diversification of cyclic nucleotide signals between CBASS and Pycsar systems is a key host adaptation to maintain successful antiphage defence^{2,11}. Acb1 and Apyc1 join a growing collection of viral nuclease enzymes dedicated to immune evasion, including phage ring nucleases that degrade cA₄ and cA₆ signals used in type III CRISPR immunity^{27,28} and poxin enzymes that degrade 2'3'-cGAMP to inhibit cyclic GMP–AMP synthase (cGAS)–stimulator of interferon genes (STING) signalling in animals²⁹. Each of these viral enzymes is structurally distinct, demonstrating at least four separate instances of prokaryotic and eukaryotic viral evolution to degrade host cyclic nucleotide immune signals. The broad specificity of Acb1 allows evasion of diverse CBASS operons with a single gene, and the ability of Acb1 to cleave cyclic trinucleotide species suggests that this enzyme may also enable evasion of type III CRISPR systems that use cAAA signals. Notably, Acb1 is unable to cleave the non-canonical 2'–5' linkage in the CBASS signalling molecule 3'2'-cGAMP (ref. 30), mirroring the recent demonstration that 3'2'-cGAMP signalling in animals enables

resistance to toxin enzymes³¹. The large diversity of >180 possible nucleotide signals proposed to exist in antiphage defence suggests that in addition to signal degradation, phages may encode Acb and Apyc proteins that target alternative components of CBASS or Pycsar immunity. Overall, our results define viral nucleases as a widespread mechanism of CBASS and Pycsar immune evasion and reveal the role of viral proteins in driving evolution of cyclic nucleotide-based immune defence systems.

Methods

Bacterial strains and phages

E. coli strain MG1655 (ATCC 47076) and *B. subtilis* BEST7003 (ref. ³²) were grown in magnesium–manganese broth (MMB; lysogeny broth supplemented with 0.1 mM MnCl₂ and 5 mM MgCl₂) with or without 0.5% agar at 37 °C or 30 °C, respectively. Whenever applicable, media were supplemented with ampicillin (100 µg ml⁻¹), chloramphenicol (34 µg ml⁻¹) or kanamycin (50 µg ml⁻¹) to ensure the maintenance of plasmids. Phage isolation was performed as previously described³³. In general, phage infections were performed in MMB media at 37 °C for *E. coli* MG1655 and at 30 °C for *B. subtilis* and phages were propagated by picking a single phage plaque into a liquid culture grown to an optical density at 600 nm (OD₆₀₀) of 0.3 in MMB medium until culture collapse. The culture was then centrifuged for 10 min at 3,200g, and the supernatant was filtered through a 0.2-µm filter. The titre of the lysate was determined using the small-drop plaque assay method as described previously³⁴.

Recombinant protein expression and purification

Acb1, Apyc1, cGAS/DncV-like nucleotidyltransferases (CD-NTase), cGAS-like receptors and effector proteins were purified from *E. coli* as previously described^{11,23,31,35}. Briefly, genes were cloned from synthetic DNA fragments (Integrated DNA Technologies) into custom pET expression vectors containing amino-terminal 6×His-SUMO2 or 6×His-MBP-SUMO2 tags by Gibson assembly using HiFi DNA Assembly Master Mix (NEB)³⁵. Expression plasmids were transformed into BL21(DE3) RIL cells (Agilent)

and plated onto MDG media (1.5% Bacto agar, 0.5% glucose, 25 mM Na₂HPO₄, 25 mM KH₂PO₄, 50 mM NH₄Cl, 5 mM Na₂SO₄, 0.25% aspartic acid, 2–50 µM trace metals, 100 µg ml⁻¹ ampicillin, 34 µg ml⁻¹ chloramphenicol). After overnight incubation at 37 °C, three colonies were used to inoculate a 30-ml MDG starter culture for 16 h (37 °C, 230 r.p.m.). M9ZB expression cultures of 1 l in volume (47.8 mM Na₂HPO₄, 22 mM KH₂PO₄, 18.7 mM NH₄Cl, 85.6 mM NaCl, 1% casamino acids, 0.5% glycerol, 2 mM MgSO₄, 2–50 µM trace metals, 100 µg ml⁻¹ ampicillin, 34 µg ml⁻¹ chloramphenicol) were then inoculated with 15 ml MDG starter culture and grown (37 °C, 230 r.p.m.) to an OD₆₀₀ of 2.5 before induction with 0.5 mM isopropyl-β-D-thiogalactoside (IPTG) for 16 h (16 °C, 230 r.p.m.). For WT Apyc1 protein, bacteria were grown in 2YT media (16 g l⁻¹ Bacto tryptone, 10 g l⁻¹ yeast extract, 5 g l⁻¹ NaCl, pH 7.0) for both starter and expression cultures and grown to an OD₆₀₀ of 1.5 before induction with 0.5 mM IPTG for 16 h (16 °C, 230 r.p.m.). Selenomethionine-labelled protein was prepared as previously described²⁹ by expressing 1 l cultures in modified M9ZB media (47.8 mM Na₂HPO₄, 22 mM KH₂PO₄, 18.7 mM NH₄Cl, 85.6 mM NaCl, 0.4% glucose, 2 mM MgSO₄, 2–50 µM trace metals, 1 µg ml⁻¹ thiamine, 100 µg ml⁻¹ ampicillin, 34 µg ml⁻¹ chloramphenicol) and allowing the cultures to grow to an OD₆₀₀ of 0.8 before supplementation with l-amino acids (50 mg ml⁻¹ leucine, isoleucine, valine; 100 mg ml⁻¹ of phenylalanine, lysine, threonine; 75 mg ml⁻¹ selenomethionine) and induction with 0.5 mM IPTG for 16 h (16 °C, 230 r.p.m.).

After overnight expression, cell pellets were collected by centrifugation and then resuspended and lysed by sonication in 50 ml lysis buffer (20 mM HEPES-KOH pH 7.5, 400 mM NaCl, 10% glycerol, 30 mM imidazole, 1 mM TCEP). Lysate was clarified by centrifugation at 50,000g for 30 min, supernatant was poured over 8 ml Ni-NTA resin (Qiagen), resin was washed with 35 ml lysis buffer supplemented with 1 M NaCl, and protein was eluted with 10 ml lysis buffer supplemented with 300 mM imidazole. Samples were then dialysed overnight in dialysis tubing with a 14 kDa molecular weight cutoff (Ward's Science), and SUMO2 tag cleavage was carried out with recombinant human SENP2 protease as previously described³⁵. Proteins used for crystallography were dialysed overnight at 4 °C in dialysis buffer

(20 mM HEPES-KOH pH 7.5, 250 mM KCl, 1 mM TCEP), and then purified further by size-exclusion chromatography using a 16/600 Superdex 75 column (Cytiva), whereas proteins used for biochemical assays were dialysed in dialysis buffer supplemented with 10% glycerol. Purified proteins were concentrated to >15 mg ml $^{-1}$ using 10-kDa MWCO centrifugal filter units (Millipore Sigma), aliquoted, flash frozen in liquid nitrogen and stored at -80 °C.

Thin-layer chromatography

Thin-layer chromatography was used to analyse cyclic nucleotide degradation as previously described²⁹. Cyclic nucleotides were synthesized using the following purified recombinant enzymes: *V. cholerae* DncV (ref. 11): cAA, 3'3'-cGAMP, cGG; *E. cloacae* CdnD (ref. 23): cAAA, cAAG; *Rhodothermus marinus* CdnE (ref. 11): cUA; *Y. aleksiciae* CdnE (ref. 36): 3'3'-cUU; *Drosophila eugracilis* cGLR1 (ref. 31): 3'2'-cGAMP; *Mus musculus* cGAS (ref. 35): 2'3'-cGAMP; *E. coli* PycC (ref. 2): cCMP; *Burkholderia cepacia* PycC (ref. 2): cUMP. Synthesis reactions were performed at 37 °C for 20 h, and consisted of 2.5 μM appropriate enzyme, 25 μM appropriate nucleoside triphosphates (NTPs), trace amounts of α -³²P-labelled NTP, 100 mM KCl, 1 mM dithiothreitol (DTT), 5 mM MgCl₂, 1 mM MnCl₂ and 50 mM Tris-HCl pH 7.5 (DncV, cGLR1, cGAS) or pH 9.0 (all other enzymes) in a final volume of 40 μl. Unincorporated NTPs were digested by addition of 1 μl Quick CIP (NEB) followed by incubation at 37 °C for 30 min and heat inactivation at 95 °C for 2 min. Synthesis reactions were then used as inputs for downstream degradation reactions, which were carried out at 37 °C in 10-μl mixtures composed of 1 μl of a 10× recombinant enzyme stock or cellular lysate, 0.25–0.5 μl of the appropriate synthesis reaction (about 1–2 μM α -³²P-labelled cyclic nucleotide), 50 mM Tris-HCl pH 7.5, 10 mM KCl and 1 mM TCEP. After 5–20-min incubation (unless indicated otherwise), 0.5 μl volumes of reactions were spotted on a 20 cm × 20 cm PEI cellulose thin-layer chromatography plate (Sigma Aldrich) and developed in 1.5 M KH₂PO₄ (pH 3.8) buffer for 45 min. Plates were dried at room temperature, exposed to a storage phosphor screen, and detected with a Typhoon Trio Variable Mode Imager System (GE Healthcare).

Cell lysate preparation

Overnight cultures of *E. coli* or *B. subtilis* were diluted 1:100 in 250 ml MMB medium and grown at 37 °C for *E. coli* and 30 °C for *B. subtilis* (250 r.p.m.) until reaching an OD₆₀₀ of 0.3. The cultures were infected with phages (Supplementary Table 1) at a final multiplicity of infection of 2. Samples of infected cells were taken before culture collapse (for time points, see Supplementary Table 1). Samples of 5 ml in volume were taken and centrifuged for 5 min at 3,200g and 4 °C. The culture pellets were flash frozen using dry ice and ethanol. *E. coli* pellets were resuspended in 250 µl of a lysis buffer containing 20 mM HEPES-KOH pH 7.5, 150 mM NaCl, 5 mM MgCl₂, 1 mM MnCl₂, 1 mM DTT, 10% glycerol and 1% NP-40, and incubated at room temperature for 30 min with occasional vortexing. *Bacillus* pellets were first treated with T4 lysozyme (ThermoFisher) at 1 mg ml⁻¹ in PBS at 37 °C for 10 min, followed by addition of 400 µl of *E. coli* lysis buffer and 30-min incubation at room temperature. Samples were clarified by centrifugation for 5 min at 17,000g at 4 °C, and the supernatant was aliquoted and flash frozen in liquid nitrogen, and stored at -80 °C.

T4 Acb1 activity-guided fractionation, mass spectrometry analysis and candidate screen

Overnight *E. coli* MG1655 cultures were diluted 1:100 into a volume of 2 l MMB and grown for about 1 h to an OD₆₀₀ of 0.3–0.5. Phage T4 was added at a multiplicity of infection of 2, and cells were collected 25 min post infection by centrifugation for 20 min at 3,200g. Infected cell pellets were resuspended in 40 ml of lysis buffer consisting of 20 mM HEPES-KOH pH 7.5, 150 mM KCl, 1 mM DTT, 5 mM MgCl₂, 1 mM MnCl₂, 10% glycerol and 1% NP-40 and incubated at room temperature for 30 min with occasional vortexing. Lysates were clarified by centrifugation at 20,000g for 15 min at 4 °C and fractionated by ion-exchange chromatography using a 5-ml HiTrap SP column and a gradient of 0.05–1.0 M NaCl. Active ion-exchange fractions were pooled, concentrated, and further separated with a 10/300 Superdex 75 column (Cytiva). In a separate approach, (NH₄)₂SO₄ was added to clarified lysates to a final concentration of 30%, and precipitated proteins were removed by centrifugation at 20,000g for 15 min.

The soluble fraction was then separated using hydrophobic interaction chromatography using a 5-ml phenyl column (Cytiva) and a gradient of 1–0.0 M $(\text{NH}_4)_2\text{SO}_4$. Active fractions were pooled, concentrated, and further separated with a 10/300 Superdex 200 column (Cytiva). For each enrichment scheme, phage T4 proteins enriched in fractions with the highest activity relative to neighbouring inactive fractions were quantified by label-free mass spectrometry as previously described²⁹.

Phage T4 genes identified by biochemical fractionation and mass spectrometry were amplified from genomic T4 DNA isolated from infected *E. coli* using a Qiagen DNeasy Blood and Tissue kit as described previously³⁷. Candidate genes were PCR amplified using Q5 DNA polymerase (NEB) and primers designed to incorporate a 49-base-pair sequence containing a T7 promoter and a ribosome-binding site upstream of the amplified candidate gene according to the NEB cell-free *E. coli* protein synthesis system instructions (NEB). PCR products were purified using a PCR clean-up kit (Qiagen) and translated using the *E. coli* protein synthesis system kit (NEB). A 1 μl volume of each translation reaction was used to test for 3'3'-cGAMP cleavage activity by thin-layer chromatography. Acb1 was identified as the product of the phage T4 gene 57B.

Phage genome sequencing, assembly and annotation of SBSphiJ1–7

SBSphiJ1–7 phages were isolated from soil samples on *B. subtilis* BEST7003 culture as described previously³³. High-titre phage lysates ($>10^7$ PFUs ml^{-1}) were used for DNA extraction. A 500 μl volume of the phage lysate was treated with DNase-I (Merck catalogue number 11284932001) added to a final concentration of 20 mg ml^{-1} and incubated at 37 °C for 1 h to remove bacterial DNA. DNA was extracted using the QIAGEN DNeasy blood and tissue kit (catalogue number 69504) starting from the Proteinase-K treatment step to lyse the phages. Libraries were prepared for Illumina sequencing using a modified Nextera protocol as previously described³⁸. Following Illumina sequencing, adapter sequences were removed from the reads using Cutadapt version 2.8 (ref. ³⁹) with the option -q 5. The trimmed reads from each phage genome were assembled into scaffolds using SPAdes

genome assembler version 3.14.0 (ref. 40), using the --careful flag. Each assembled genome was analysed with Prodigal version 2.6.3 (ref. 41; default parameters) to predict open reading frames.

SBSphiJ Apyc1 bioinformatic identification

The genomic sequences of SBSphiJ and the closely related family members SBSphiJ1–7 were aligned using progressive Mauve (ref. 42). Regions that were exclusive to cCMP-cleaving phages revealed eight candidate genes. The corresponding SBSphiJ protein sequences were analysed using HHpred (ref. 43) for predicted structural homologues. Protein classes with >75% probability are listed in Extended Data Fig. 5b and Apyc1 was identified as the product of the phage SBSphiJ gene 147.

Identification of Acb1 and Apyc1 homologues and generation of phylogenetic trees

Homologues of Acb1 and Apyc1 were identified using NCBI BLASTp with default parameters. Acb1 sequences were classified as belonging to a prophage if they were within three genes of a phage structural or packaging protein. Apyc1 phage sequences were identified by restricting the search to only viral sequences (NCBI taxid:

10293; <https://www.ncbi.nlm.nih.gov/Taxonomy/Browser/wwwtax.cgi?id=10239>). Maximum-likelihood trees were generated using the IQ-TREE web server with ultrafast bootstrapping and 1,000 iterations⁴⁴. Consensus trees were then edited visually using the Interactive Tree Of Life⁴⁵.

Crystallization and structure determination

Crystals were grown in hanging-drop format using EasyXtal 15-well trays (NeXtal). Crystals of native and selenomethionine-labelled phage FBB1 Acb1 G8–D152 were grown at 18 °C in 2-μl drops mixed 1:1 with purified protein (4 mg ml⁻¹, 20 mM HEPES-KOH pH 7.5, 80 mM KCl, 1 mM TCEP) and reservoir solution (2 M ammonium sulfate, 0.1 M sodium citrate pH 4.6). Crystals were grown for 1–7 days before being cryo-protected with reservoir solution supplemented with 45% sucrose and collected by freezing

in liquid nitrogen. Crystals of the FBB1 Acb1–3'3'-cGAMP complex were grown using the same reservoir conditions, except drops and cryo-protectant solution were supplemented with 100 μM of a hydrolysis-resistant phosphorothioate-modified analogue of 3'3'-cGAMP (Biolog Life Science Institute, C 216). Crystals of Bsp38 Apyc1 were grown at 18 °C in 2-μl drops mixed 1:1 with purified protein (10 mg ml⁻¹, 20 mM HEPES-KOH pH 7.5, 80 mM KCl, 1 mM TCEP) and reservoir solution (0.2 M lithium sulfate, 0.1 M Tris-HCl pH 7.5, 30% PEG-4000). Crystals were grown for 1–7 days before being cryo-protected with reservoir solution supplemented with 15% glycerol and collected by freezing in liquid nitrogen. Crystals of selenomethionine-labelled *Paenibacillus J14* (GenBank accession number [WP_028539944.1](#)) Apyc1 were grown at 18 °C in 2-μl drops mixed 1:1 with purified protein (10 mg ml⁻¹, 20 mM HEPES-KOH pH 7.5, 80 mM KCl, 1 mM TCEP) and reservoir solution (0.1 M Tris-HCl pH 8.5, 0.2 M MgCl₂, 16% PEG-4000) supplemented with 100 μM of a hydrolysis-resistant phosphorothioate-modified analogue of cAMP (Biolog Life Science Institute, A 003). Crystals were grown for 1–7 days before being cryo-protected with reservoir solution supplemented with 25% ethylene glycol and collected by freezing in liquid nitrogen. Crystals of *Paenibacillus xerothermodurans* Apyc1 were grown at 18 °C in 2-μl drops mixed 1:1 with purified protein (10 mg ml⁻¹, 20 mM HEPES-KOH pH 7.5, 80 mM KCl, 1 mM TCEP) and reservoir (0.1 M HEPES-KOH pH 7.5, 0.2 M calcium acetate, 10% PEG-8000). Crystals were grown for 1–7 days before being cryo-protected with reservoir solution supplemented with 25% ethylene glycol and collected by freezing in liquid nitrogen. X-ray diffraction data were collected at the Advanced Photon Source (beamlines 24-ID-C and 24-ID-E), and data were processed using the SSRL autoxds script (A. Gonzalez, Stanford SSRL). For Acb1 and Apyc1 phase determination, anomalous data were collected using selenomethionine-labelled Acb1 crystals, heavy sites were identified with HySS in Phenix (ref. [46](#)), and an initial map was produced using SOLVE/RESOLVE in Phenix (ref. [46](#)). Model building was performed using Coot (ref. [47](#)), and then refined in Phenix. Statistics were analysed as described in Supplementary Table [2](#) (refs. [48,49,50](#)). Final structures were refined to stereochemistry statistics for Ramachandran plot (favoured/allowed), rotamer outliers and MolProbity score as follows: FBB1 Acb1, 98.52%/1.48%, 0.8% and 1.11; FBB1 Acb1–3'3'-cGAMP, 99.26%/0.74%, 1.56% and 1.39; Bsp38 Apyc1, 90.79%/7.46%, 4.85% and

2.64; *P. J14* Apyc1, 95.04%/4.96%, 2.38% and 1.78; *P. xerothermodurans* Apyc1, 96.12%/3.88%, 1.93% and 1.60. See Supplementary Table 2 and the Data availability statement for the deposited PDB codes. All structure figures were generated with PyMOL 2.3.0.

HPLC

Acb1 and Apyc1 reactions for HPLC analysis were performed in a 100 µl volume and consisted of 50 mM Tris-HCl pH 7.5, 100 mM KCl, 1 mM DTT, 100 µM chemically synthesized nucleotide standards (Biolog Life Science Institute) and 1 µM recombinant protein unless otherwise indicated. Apyc1 reactions were further supplemented with 5 mM MgCl₂ and 1 mM MnCl₂. Reactions were incubated at 37 °C for 20 min (unless otherwise indicated in the figure legend) and filtered using a 10-kDa cutoff filter (Millipore). Filtered nucleotide products were analysed using a C18 column (Agilent Zorbax Bonus-RP 4.6 × 150 mm, 3.5 µm) heated to 40 °C and run at 1 ml min⁻¹ in a buffer of 50 mM NaH₂PO₄ adjusted to pH 6.8 with NaOH, supplemented with 3% acetonitrile.

In vitro reconstitution of CBASS and Pycsar effector function and inhibition

Synthetic cyclic nucleotides (Biolog Life Science Institute) were pre-incubated with purified T4 Acb1 and SBSphiJ Apyc1 in reactions containing 1 µM cyclic nucleotide, 1 µM recombinant Acb1 or Apyc1 protein, 50 mM Tris-HCl pH 7.5, 100 mM KCl and 1 mM DTT for 1 h at 37 °C. Apyc1 reactions were further supplemented with 5 mM MgCl₂ and 1 mM MnCl₂. Cyclic nucleotide reactions were then used as 10× inputs for effector activation reactions using the following recombinant CBASS and Pycsar effector proteins: *V. cholerae* CapV (ref. 11), *E. cloacae* Cap4 (ref. 23), *B. pseudomallei* Cap5 (ref. 23) and *B. cepacia* PycTIR (ref. 2). Nuclease effectors were incubated in 25-µl reactions containing 1 µM effector protein, and buffer consisting of 50 mM Tris-HCl pH 7.5, 25 mM NaCl, 5 mM MgCl₂, 1 mM DTT and 10 ng µl⁻¹ pGEM9z plasmid DNA. Following 20-min incubation at 37 °C, 5 µl of DNA loading dye was added and 15 µl was analysed on a 1% agarose gel as previously described²³. CapV

phospholipase activity was analysed in 25- μ l reactions consisting of 1 μ M purified effector, 50 mM Tris-HCl pH 7.5, 25 mM NaCl, 5 mM MgCl₂, 1 mM DTT and a BODIPY-labelled EnzChek phospholipase substrate (ThermoFisher) as previously described¹¹. Phospholipase activity was measured using a Synergy H1 plate reader (BioTek) according to the manufacturer's instructions. PycTIR was used at 40 μ M in 25- μ l reactions consisting of 20 mM HEPES-KOH pH 7.5, 100 mM KCl and 500 μ M of the fluorescent NAD⁺ analogue ε -NAD (Sigma). Fluorescent measurements (300 nm excitation, 410 nm emission) were taken in a Synergy H1 plate reader (BioTek) following 2-min incubation at room temperature.

Bacterial growth assays

CBASS effector function was measured in *E. coli* using conditions that result in autoactivation of *V. cholerae* DncV 3'3'-cGAMP synthesis as previously described³⁶. *E. coli* BL21(DE3) competent cells (NEB) were transformed with three plasmids encoding *V. cholerae* CapV (pBAD33), the CBASS effector *B. pseudomallei* Cap5 (pET16), and either WT or catalytically inactive (H44A/H113A) T4 Acb1 (pTU175)⁵¹. Transformations were plated onto MDG plates and three colonies were picked and grown for 16 h (37 °C, 230 r.p.m.) in 5-ml MDG starter cultures. A 5 ml volume of M9ZB cultures was inoculated with 200 μ l MDG starter culture and grown for 3 h (37 °C, 230 r.p.m.) before being induced by diluting 1:5 in M9ZB media containing 5 μ M IPTG and 0.2% l-arabinose. Induced culture (200 μ l) was added to wells of a 96-well plate, and OD₆₀₀ was read every 6.82 min for 300 min in a Synergy H1 plate reader (BioTek) while shaking at 230 r.p.m., 37 °C. Wells containing medium alone were used for OD₆₀₀ background subtraction.

Phage challenge assays

Phage challenge experiments were performed as previously described^{1,2} by spotting serial dilutions of high-titre phage stocks onto a lawn of bacteria carrying a complete CBASS or Pycsar defence operon. The following defence systems were used: *E. coli* strain KTE188 (IMG gene accession numbers: 2564596481–2564596485; <https://img.jgi.doe.gov/>) cloned under

its native promoter into the plasmid pSG1 (ref. 3), *E. coli* CdnG cloned under its native promoter into the plasmid pLOCO2 (ref. 23), *Y. aleksiciae* CdnE (ref. 36) cloned into a pBAD vector, and *E. coli* PycC (ref. 2) cloned under its native promoter into the plasmid pSG1. For *EcCdnG* and *YaCdnE* operons, control plasmids were also used in which the CD-NTase is inactivated (*CdnG*-D82A/D84A)²³ or the transmembrane segment of the receptor is deleted (*YaCdnE*)³⁶. Phage replication in the context of these defence systems was measured using a spot plaque assay³⁶. Briefly, *E. coli* MG1655 (*EcKTE188*, *EcPycC*) or BL21 cells (*EcCdnG* and *YaCdnE*) containing the defence systems were grown overnight at 37 °C. A 300 µl volume of the bacterial culture was mixed with 4 ml melted MMB agar containing appropriate antibiotics and 0.2% arabinose for pBAD plasmids, poured on top of a 15-cm plate of lysogeny broth and left to solidify in a plate for 1 h at room temperature. High-titre phage stocks were serially diluted tenfold in MMB and 3–5-µl drops were placed on the bacterial layer and allowed to dry at room temperature for 1 h. Plates were incubated overnight at 37 °C (Acb1 and Apyc1 rescue experiments) or 30 °C ($\Delta acb1$ T4 phage challenges) and plaque-forming units (PFUs) were determined by counting the derived plaques after overnight incubation. Phage infection of cells expressing active CBASS operons did not generate clear plaques. For these, the dilution at which there was no detectable defect in bacterial growth was counted as having a single plaque. For in vivo rescue experiments, *acb1* and *apyc1* were amplified from the genome of T4 phage or SBSphiJ phage and cloned into the plasmid pBbS8k (Addgene number 35276) using Gibson assembly (NEB).

Generation of phage T4 $\Delta Acb1$

Nonsense mutations were introduced into *acb1* using a CRISPR-based selection strategy as described previously^{52,53}. Briefly, a gRNA targeting *acb1* and a repair template with nonsense mutations were cloned into pCRISPR (Addgene 42875). *E. coli* Top10 cells were then transformed with the pCRISPR–gRNA-*acb1* repair plasmid and pCas9 (Addgene 42876). A colony was picked, and 2-ml log-scale cultures were infected with WT phage T4 until culture collapse. The resulting lysate was filtered through a 0.22-µM filter and plated on *E. coli* Top10 cells with no plasmid. Single plaques were picked into 200 µl SM buffer (50 mM Tris-HCl pH 8.5, 100

mM NaCl, 8 mM MgSO₄) containing 2 µl chloroform. After 1-h incubation at room temperature, 4 µl was used as input for standard PCR reactions using GoTaqGreen (Promega) according to the manufacturer's instructions. PCR products were purified using QIAquick gel extraction kit (Qiagen) and sequenced for introduction of nonsense mutations. Positive phage T4 clones went through three rounds of plaque purification before generating a high-titre stock used in all phage challenge experiments.

Statistics and reproducibility

Statistical tests are described in the figure legends and were performed using GraphPad Prism 9.3.1. Experimental details regarding replicates and sample size are described in the figure legends. No statistical methods were used to predetermine sample size and no blinding or randomization was used for this study.

Reporting summary

Further information on research design is available in the [Nature Research Reporting Summary](#) linked to this paper.

Data availability

Coordinates and structure factors of FBB1 Acb1, the FBB1 Acb1–3'3'-cGAMP complex, Bsp38 Apyc1, *P. J14* Apyc1 and *P. xerothermodurans* Apyc1 have been deposited in the PDB under the accession codes [7T26](#), [7T27](#), [7T28](#), [7U2R](#) and [7U2S](#), respectively. Source data are provided with this paper. All other data are available in the manuscript or the [Supplementary Information](#).

References

1. Cohen, D. et al. Cyclic GMP–AMP signalling protects bacteria against viral infection. *Nature* **574**, 691–695 (2019).

2. Tal, N. et al. Cyclic CMP and cyclic UMP mediate bacterial immunity against phages. *Cell* **184**, 5728–5739 (2021).
3. Wiegand, T., Karambelkar, S., Bondy-Denomy, J. & Wiedenheft, B. Structures and strategies of anti-CRISPR-mediated immune suppression. *Annu. Rev. Microbiol.* **74**, 21–37 (2020).
4. Bickle, T. A. & Kruger, D. H. Biology of DNA restriction. *Microbiol. Rev.* **57**, 434–450 (1993).
5. Meeske, A. J. et al. A phage-encoded anti-CRISPR enables complete evasion of type VI-A CRISPR-Cas immunity. *Science* **369**, 54–59 (2020).
6. Jia, N. & Patel, D. J. Structure-based functional mechanisms and biotechnology applications of anti-CRISPR proteins. *Nat. Rev. Mol. Cell Biol.* **22**, 563–579 (2021).
7. Bondy-Denomy, J., Pawluk, A., Maxwell, K. L. & Davidson, A. R. Bacteriophage genes that inactivate the CRISPR/Cas bacterial immune system. *Nature* **493**, 429–432 (2013).
8. Hampton, H. G., Watson, B. N. J. & Fineran, P. C. The arms race between bacteria and their phage foes. *Nature* **577**, 327–336 (2020).
9. Chowdhury, S. et al. Structure reveals mechanisms of viral suppressors that intercept a CRISPR RNA-guided surveillance complex. *Cell* **169**, 47–57 (2017).
10. Stanley, S. Y. & Maxwell, K. L. Phage-encoded anti-CRISPR defenses. *Annu. Rev. Genet.* **52**, 445–464 (2018).
11. Whiteley, A. T. et al. Bacterial cGAS-like enzymes synthesize diverse nucleotide signals. *Nature* **567**, 194–199 (2019).
12. Jenal, U., Reinders, A. & Lori, C. Cyclic di-GMP: second messenger extraordinaire. *Nat. Rev. Microbiol.* **15**, 271–284 (2017).

13. Stulke, J. & Kruger, L. Cyclic di-AMP signaling in bacteria. *Annu. Rev. Microbiol.* **74**, 159–179 (2020).
14. Pinilla-Redondo, R. et al. Discovery of multiple anti-CRISPRs highlights anti-defense gene clustering in mobile genetic elements. *Nat. Commun.* **11**, 5652 (2020).
15. Bair, C. L., Rifat, D. & Black, L. W. Exclusion of glucosyl-hydroxymethylcytosine DNA containing bacteriophages is overcome by the injected protein inhibitor IPI*. *J. Mol. Biol.* **366**, 779–789 (2007).
16. Kazlauskiene, M., Kostiuk, G., Venclovas, C., Tamulaitis, G. & Siksnys, V. A cyclic oligonucleotide signaling pathway in type III CRISPR-Cas systems. *Science* **357**, 605–609 (2017).
17. Niewoehner, O. et al. Type III CRISPR–Cas systems produce cyclic oligoadenylate second messengers. *Nature* **548**, 543–548 (2017).
18. Shi, H., Wu, J., Chen, Z. J. & Chen, C. Molecular basis for the specific recognition of the metazoan cyclic GMP-AMP by the innate immune adaptor protein STING. *Proc. Natl Acad. Sci. USA* **112**, 8947–8952 (2015).
19. Kranzusch, P. J. et al. Ancient origin of cGAS-STING reveals mechanism of universal 2',3' cGAMP signaling. *Mol. Cell* **59**, 891–903 (2015).
20. Richter, W. 3',5' Cyclic nucleotide phosphodiesterases class III: members, structure, and catalytic mechanism. *Proteins* **46**, 278–286 (2002).
21. Tian, Y. et al. Dual specificity and novel structural folding of yeast phosphodiesterase-1 for hydrolysis of second messengers cyclic adenosine and guanosine 3',5'-monophosphate. *Biochemistry* **53**, 4938–4945 (2014).
22. Na, H. W., Namgung, B., Song, W. S. & Yoon, S. I. Structural and biochemical analyses of the metallo-beta-lactamase fold protein YhfI

- from *Bacillus subtilis*. *Biochem. Biophys. Res. Commun.* **519**, 35–40 (2019).
23. Lowey, B. et al. CBASS immunity uses CARF-related effectors to sense 3'-5'- and 2'-5'-linked cyclic oligonucleotide signals and protect bacteria from phage infection. *Cell* **182**, 38–49 (2020).
 24. Lau, R. K. et al. Structure and mechanism of a cyclic trinucleotide-activated bacterial endonuclease mediating bacteriophage immunity. *Mol. Cell* **77**, 723–733 (2020).
 25. Severin, G. B. et al. Direct activation of a phospholipase by cyclic GMP-AMP in El Tor *Vibrio cholerae*. *Proc. Natl Acad. Sci. USA* **115**, E6048–E6055 (2018).
 26. Morehouse, B. R. et al. STING cyclic dinucleotide sensing originated in bacteria. *Nature* **586**, 429–433 (2020).
 27. Athukoralage, J. S. et al. An anti-CRISPR viral ring nuclease subverts type III CRISPR immunity. *Nature* **577**, 572–575 (2020).
 28. Athukoralage, J. S. et al. The dynamic interplay of host and viral enzymes in type III CRISPR-mediated cyclic nucleotide signalling. *eLife* **9**, e55852 (2020).
 29. Eaglesham, J. B., Pan, Y., Kupper, T. S. & Krantzsch, P. J. Viral and metazoan poxins are cGAMP-specific nucleases that restrict cGAS-STING signalling. *Nature* **566**, 259–263 (2019).
 30. Fatma, S., Chakravarti, A., Zeng, X. & Huang, R. H. Molecular mechanisms of the CdnG-Cap5 antiphage defense system employing 3',2'-cGAMP as the second messenger. *Nat. Commun.* **12**, 6381 (2021).
 31. Slavik, K. M. et al. cGAS-like receptors sense RNA and control 3'2'-cGAMP signalling in *Drosophila*. *Nature* **597**, 109–113 (2021).
 32. Barrangou, R. & van der Oost, J. Bacteriophage exclusion, a new defense system. *EMBO J.* **34**, 134–135 (2015).

33. Doron, S. et al. Systematic discovery of antiphage defense systems in the microbial pangenome. *Science* **359**, eaar4120 (2018).
34. Mazzocco, A., Waddell, T. E., Lingohr, E. & Johnson, R. P. Enumeration of bacteriophages using the small drop plaque assay system. *Methods Mol. Biol.* **501**, 81–85 (2009).
35. Zhou, W. et al. Structure of the human cGAS-DNA complex reveals enhanced control of immune surveillance. *Cell* **174**, 300–311 (2018).
36. Duncan-Lowey, B., McNamara-Bordewick, N. K., Tal, N., Sorek, R. & Kranzusch, P. J. Effector-mediated membrane disruption controls cell death in CBASS antiphage defense. *Mol. Cell* **81**, 5039–5051 (2021).
37. Jakociune, D. & Moodley, A. A rapid bacteriophage DNA extraction method. *Methods Protoc.* **1**, 27 (2018).
38. Baym, M. et al. Inexpensive multiplexed library preparation for megabase-sized genomes. *PLoS ONE* **10**, e0128036 (2015).
39. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J.* <https://doi.org/10.14806/ej.17.1.200> (2011).
40. Nurk, S. et al. Assembling single-cell genomes and mini-metagenomes from chimeric MDA products. *J. Comput. Biol.* **20**, 714–737 (2013).
41. Hyatt, D. et al. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinform.* **11**, 119 (2010).
42. Darling, A. E., Mau, B. & Perna, N. T. progressiveMauve: multiple genome alignment with gene gain, loss and rearrangement. *PLoS ONE* **5**, e11147 (2010).
43. Zimmermann, L. et al. A completely reimplemented MPI bioinformatics toolkit with a new HHpred server at its core. *J. Mol. Biol.* **430**, 2237–2243 (2018).

44. Trifinopoulos, J., Nguyen, L. T., von Haeseler, A. & Minh, B. Q. W-IQ-TREE: a fast online phylogenetic tool for maximum likelihood analysis. *Nucleic Acids Res.* **44**, W232–W235 (2016).
45. 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).
46. Liebschner, D. et al. Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix. *Acta Crystallogr. D* **75**, 861–877 (2019).
47. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D* **60**, 2126–2132 (2004).
48. Chen, V. B. et al. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D* **66**, 12–21 (2010).
49. Karplus, P. A. & Diederichs, K. Linking crystallographic model and data quality. *Science* **336**, 1030–1033 (2012).
50. Weiss, M. S. Global indicators of X-ray data quality. *J. Appl. Cryst.* **34**, 130–135 (2001).
51. Uehara, T., Parzych, K. R., Dinh, T. & Bernhardt, T. G. Daughter cell separation is controlled by cytokinetic ring-activated cell wall hydrolysis. *EMBO J.* **29**, 1412–1422 (2010).
52. Tao, P., Wu, X., Tang, W. C., Zhu, J. & Rao, V. Engineering of bacteriophage T4 genome using CRISPR-Cas9. *ACS Synth. Biol.* **6**, 1952–1961 (2017).
53. Duong, M. M., Carmody, C. M., Ma, Q., Peters, J. E. & Nugen, S. R. Optimization of T4 phage engineering via CRISPR/Cas9. *Sci. Rep.* **10**, 18229 (2020).

Acknowledgements

We are grateful to B. Duncan-Lowey, and members of the laboratory of P.J.K. for helpful comments and discussion. Mass spectrometry was performed at the Biopolymers and Proteomics Core Facility at the Koch Institute of the Massachusetts Institute of Technology and the Taplin Mass Spectrometry Facility at Harvard Medical School. The work was financed by grants to P.J.K. from the Pew Biomedical Scholars programme, the Burroughs Wellcome Fund PATH programme, The Mathers Foundation, The Mark Foundation for Cancer Research, the Parker Institute for Cancer Immunotherapy and the National Institutes of Health (1DP2GM146250-01) and grants to R.S. from the European Research Council (ERC-AdG GA 101018520), the Israel Science Foundation (ISF 296/21), the Ernest and Bonnie Beutler Research Program of Excellence in Genomic Medicine, the Deutsche Forschungsgemeinschaft (SPP 2330, grant 464312965), the Minerva Foundation and the Knell Family Center for Microbiology. S.J.H. is supported through a Cancer Research Institute Irvington Postdoctoral Fellowship (CRI3996), T.W. is supported through a Minerva Foundation postdoctoral fellowship and a European Molecular Biology Organization postdoctoral fellowship (ALTF 946-2020), and B.R.M. is supported as a Ruth L. Kirschstein National Research Service Award Postdoctoral Fellow (NIH F32GM133063). X-ray data were collected at the Northeastern Collaborative Access Team beamlines 24-ID-C and 24-ID-E (P30 GM124165), and used a Pilatus detector (S10RR029205), an Eiger detector (S10OD021527) and the Argonne National Laboratory Advanced Photon Source (DE-AC02-06CH11357).

Author information

Affiliations

1. Department of Microbiology, Harvard Medical School, Boston, MA, USA

Samuel J. Hobbs, Allen Lu, Benjamin R. Morehouse, Julia Schnabel & Philip J. Kranzusch

2. Department of Cancer Immunology and Virology, Dana-Farber Cancer Institute, Boston, MA, USA

Samuel J. Hobbs, Allen Lu, Benjamin R. Morehouse, Julia Schnabel & Philip J. Kranzusch

3. Department of Molecular Genetics, Weizmann Institute of Science, Rehovot, Israel

Tanita Wein, Azita Leavitt, Erez Yirmiya & Rotem Sorek

4. Parker Institute for Cancer Immunotherapy at Dana-Farber Cancer Institute, Boston, MA, USA

Philip J. Kranzusch

Contributions

Experiments were designed and conceived by S.J.H. and P.J.K. Anti-CBASS and anti-Pycsar biochemical screen was performed by S.J.H. with phage-infected samples prepared by T.W. Phage Acb1 and Apyc1 identification and validation experiments were performed by S.J.H. Crystallography and biochemical experiments were performed by S.J.H. and A. Lu, with assistance from J.S. NADase effector assays were performed by B.R.M. Phage genome sequencing was performed by A. Leavitt and E.Y. Phage challenge assays were performed by T.W., S.J.H. and R.S. The manuscript was written by S.J.H. and P.J.K. All authors contributed to editing the manuscript and support the conclusions.

Corresponding author

Correspondence to [Philip J. Kranzusch](#).

Ethics declarations

Competing interests

R.S. is a scientific cofounder and advisor of BiomX and Ecophage. 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 A biochemical screen to discover coliphage anti-CBASS and anti-Pycsar evasion.](#)

a, Representative TLC assays depicting coliphage degradation of radiolabeled cyclic nucleotides after 1 h incubation in infected *E. coli* lysates.

[Extended Data Fig. 2 A biochemical screen to discover *Bacillus* phage anti-CBASS and anti-Pycsar evasion.](#)

a, Representative TLC assays depicting *Bacillus* phage degradation of radiolabeled cyclic nucleotides after 1 h incubation in infected *B. subtilis* lysates.

[Extended Data Fig. 3 Biochemical fractionation and mass spectrometry identification of phage T4 gene 57B as Acb1.](#)

a, Schematic depicting strategies of biochemical fractionation to enrich 3'3'-cGAMP cleavage activity from crude T4 lysate. Mass spectrometry of active fractions revealed 37 candidate T4 proteins. **b**, Agarose gel analysis of 34 successfully PCR-amplified candidate T4 genes to be screened for 3'3'-

cGAMP cleavage activity by *in vitro* translation. Data are representative of 2 independent experiments. For gel source data, see Supplementary Figure 1. **c**, Translation products from b were tested for 3'3'-cGAMP cleavage activity by TLC. Data are representative of 2 independent experiments.

Extended Data Fig. 4 Purification and biochemical characterization of recombinant T4 Acb1.

a, T4 Acb1 was expressed as an N-terminal 6×His-MBP-SUMO2 fusion and purified by Ni-NTA and separated from His-MBP-SUMO2 by size exclusion chromatography. **b**, Coomassie-stained SDS-PAGE analysis of fully purified T4 Acb1. **c**, TLC analysis of 3'3'-cGAMP degradation by T4 Acb1 at the indicated pH. **d**, TLC analysis of 3'3'-cGAMP degradation by T4 Acb1 supplemented with the indicated metal or EDTA at the following concentrations: 50 mM EDTA; 5 mM MgCl₂; 1 mM MnCl₂; 5 mM CaCl₂; 1 μM NiCl₂; 1 μM CuCl₂; or 1 μM ZnSO₄. Data in all panels are representative of at least 3 independent experiments.

Extended Data Fig. 5 Bioinformatic identification and biochemical characterization of phage SBSphiJ gene 147 as the anti-Pycsar nuclease Apyc1.

a, Genome schematic of SBSphiJ and 7 other closely related phages highlighting regions exclusive to cCMP-cleaving phages. **b**, Summary of HHpred analysis of candidate genes. **c**, Recombinant SBSphiJ Apyc1 was expressed as an N-terminal 6×His-SUMO2 fusion, purified by Ni-NTA, and separated from His-SUMO2 by size exclusion chromatography. **d**, Coomassie-stained SDS-PAGE analysis of fully purified SBSphiJ Apyc1. **e**, TLC analysis of cCMP degradation by SBSphiJ Apyc1 at the indicated pH. **f**, HPLC analysis of recombinant SBSphiJ Apyc1 incubated with the indicated substrates (100 μM) for 30 min at 37 °C. Data in c–f are representative of 3 independent experiments.

Extended Data Fig. 6 Substrate specificity of host and viral enzymes related to Acb1 and Apyc1.

a, Homologs of T4 Acb1 were expressed, purified, and tested for cleavage of 3'3'-cGAMP (left) and cAAA (right) by TLC. Data are representative of 2 independent experiments. **b**, Summary of the distribution of bacterial Apyc1 homologs among bacterial orders. **c**, Homologs of SBSphiJ Apyc1 were expressed, purified, and tested for cleavage of cAMP and cCMP by HPLC. Data are representative of 3 independent experiments. **d**, Summary of HPLC analysis of cNMP degradation by SBSphiJ Apyc1 and closely related *B. subtilis* MBL phosphodiesterases. Data are presented as mean ± s.d. from n = 3 independent replicates.

[Source Data](#)

Extended Data Fig. 7 Structural analysis of Acb1 and mechanism of 3'3'-cGAMP cleavage.

a, Structure guided multiple sequence alignment of Acb1 proteins from the indicated phages. The strength of shading indicates degree of residue conservation. **b**, Overview of Acb1 in the *apo* state (blue) and bound to 3'3'-cGAMP (green). The C-terminal lid is unstructured in the apo state and encloses 3'3'-cGAMP upon binding. **c**, Summary of HPLC analysis of WT or E141A T4 Acb1 cleavage of the indicated substrate. Data are presented as mean ± s.d. from n = 3 independent experiments. **d**, HPLC analysis of cA₄ cleavage by T4 Acb1. Data in graph are presented as mean ± s.d. from n = 3 independent experiments. **e**, Polder omit map of 3'3'-cGAMP contoured at 3.0 σ. **f**, Comparison of T4 Acb1 3'3'-cGAMP degradation products and synthetic 3'3'-cGAMP and GpAp standards by HPLC. Data are representative of 2 independent experiments.

[Source Data](#)

Extended Data Fig. 8 Structural analysis of Apyc1 and mechanism of cNMP degradation.

a, Structure guided multiple sequence alignment of Apyc1 proteins from the indicated phages or bacterial species and *B. subtilis* MBL phosphodiesterases. The strength of shading indicates degree of residue conservation. **b**, Overview of *P. J14* Apyc1, *P. xerothermodurans* Apyc1,

Bsp38 Apyc1, and *B. subtilis* YhfI crystal structures, with one monomer in colour and one monomer in grey. Detailed area highlights an Apyc1-specific loop that extends into the cNMP binding pocket. **c**, Detailed view of the residues coordinating the Zn²⁺ ions with cAMP modeled into the cNMP binding pocket. Numbers in parentheses indicate equivalent residue in SBSphiJ Apyc1. **d**, *P. J14* Apyc1 crystallized in the presence of a hydrolysis-resistant phosphorothioate analog of cAMP resulted in clear phosphate and ribose density in the binding pocket and sparse density corresponding to the nucleobase. Polder omit map of cAMP contoured at 3.0 σ. **e**, TLC analysis of cCMP cleavage by SBSphiJ Apyc1 point mutants. **f**, HPLC analysis of Apyc1 cCMP and cUMP reaction products compared to synthesized 5'-CMP, 5'-UMP and 3'-CMP or 3'-UMP standards. Data are representative of 2 independent experiments.

Extended Data Fig. 9 Effector inhibition and time course analysis of Acb1 and Apyc1 activity.

a, Agarose gel analysis of uncut plasmid DNA incubated with Cap4 and cAAG that was treated with WT Acb1, catalytically inactive Acb1 H44A, or WT Apyc1. Data are representative of 3 independent experiments. For gel source data, see Supplementary Figure 1. **b**, Release of fluorescent dye from a phospholipid substrate incubated with recombinant CapV and 3'3'-cGAMP that was treated with WT or catalytically inactive H44A/H113A Acb1. Data are presented as mean ± s.d. from n = 3 independent experiments. **c**, T4-infected cells were collected at the indicated time point, and lysates were tested for 3'3'-cGAMP cleavage activity by TLC. Data are representative of 3 independent experiments. **d**, SBSphiJ-infected cells were collected at the indicated time point and lysates were tested for cCMP cleavage activity. Data are representative of 2 independent experiments. **e**, Bacterial growth in cells expressing Cap5, WT or D132A/D134A catalytically inactive DncV, and WT or H44A/H113A catalytically inactive T4 Acb1. Technical replicates are plotted and the data are representative of 3 independent experiments.

Source Data

Extended Data Fig. 10 Generation and validation of phage T4 Δacb1.

a, Sequencing reads of WT and $\Delta acb1$ phage T4. **b**, TLC analysis of cyclic nucleotide cleavage by WT or $\Delta acb1$ phage T4 lysate. Data are representative of 2 independent experiments. **c**, Representative plaque assays of *E. coli* carrying a plasmid encoding an active or catalytically inactive CBASS operon from *Y. aleksiciae* **d**, Summary of plaque assay results of WT or $\Delta acb1$ phage T4 infection of *E. coli* carrying catalytically inactive CBASS operons from *Y. aleksiciae* or *E. coli*. Data are presented as mean \pm s.d. from n = 4 (*Y. aleksiciae* operon) or n = 3 technical replicates (*E. coli* operon) and are representative of at least 3 biologically independent experiments.

[Source Data](#)

Supplementary information

[Supplementary Information](#)

This file contains Supplementary Fig. 1 and Tables 1 and 2.

[Reporting Summary](#)

[Peer Review File](#)

Source data

[Source Data Fig. 2](#)

[Source Data Fig. 4](#)

[Source Data Extended Data Fig. 6](#)

[Source Data Extended Data Fig. 7](#)

[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

Hobbs, S.J., Wein, T., Lu, A. *et al.* Phage anti-CBASS and anti-Pycsar nucleases subvert bacterial immunity. *Nature* **605**, 522–526 (2022). <https://doi.org/10.1038/s41586-022-04716-y>

- Received: 06 December 2021
- Accepted: 01 April 2022
- Published: 08 April 2022
- Issue Date: 19 May 2022

- DOI: <https://doi.org/10.1038/s41586-022-04716-y>

Share this article

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

[Get shareable link](#)

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

[Copy to clipboard](#)

Provided by the Springer Nature SharedIt content-sharing initiative

Further reading

- [**Killing the messenger to evade bacterial defences**](#)

◦ Urs Jenal

Nature (2022)

[**Killing the messenger to evade bacterial defences**](#)

- Urs Jenal

Nature News & Views 29 Apr 2022

This article was downloaded by **calibre** from <https://www.nature.com/articles/s41586-022-04716-y>.

[Download PDF](#)

- Article
- Open Access
- [Published: 04 May 2022](#)

Nonlinear mechanics of human mitotic chromosomes

- [Anna E. C. Meijering^{1 na1}](#),
- [Kata Sarlós^{2 na1}](#),
- [Christian F. Nielsen^{2 na1}](#),
- [Hannes Witt](#) [ORCID: orcid.org/0000-0001-8785-8293¹](#),
- [Janni Harju](#) [ORCID: orcid.org/0000-0001-9696-5453¹](#),
- [Emma Kerklingh¹](#),
- [Guus H. Haasnoot¹](#),
- [Anna H. Bizard](#) [ORCID: orcid.org/0000-0003-1858-3999²](#),
- [Iddo Heller](#) [ORCID: orcid.org/0000-0003-4888-2244¹](#),
- [Chase P. Broedersz](#) [ORCID: orcid.org/0000-0001-7283-3704¹](#),
- [Ying Liu](#) [ORCID: orcid.org/0000-0003-3703-2498²](#),
- [Erwin J. G. Peterman](#) [ORCID: orcid.org/0000-0003-1058-249X^{1 na2}](#),
- [Ian D. Hickson](#) [ORCID: orcid.org/0000-0002-0583-566X^{2 na2}](#) &
- [Gijs J. L. Wuite](#) [ORCID: orcid.org/0000-0002-5706-043X^{1 na2}](#)

[Nature](#) volume 605, pages 545–550 (2022)

- 11k Accesses
- 54 Altmetric
- [Metrics details](#)

Subjects

- [Biological fluorescence](#)
- [Single-molecule biophysics](#)
- [Supramolecular assembly](#)
- [Thermodynamics](#)

Abstract

In preparation for mitotic cell division, the nuclear DNA of human cells is compacted into individualized, X-shaped chromosomes¹. This metamorphosis is driven mainly by the combined action of condensins and topoisomerase IIα (TOP2A)^{2,3}, and has been observed using microscopy for over a century. Nevertheless, very little is known about the structural organization of a mitotic chromosome. Here we introduce a workflow to interrogate the organization of human chromosomes based on optical trapping and manipulation. This allows high-resolution force measurements and fluorescence visualization of native metaphase chromosomes to be conducted under tightly controlled experimental conditions. We have used this method to extensively characterize chromosome mechanics and structure. Notably, we find that under increasing mechanical load, chromosomes exhibit nonlinear stiffening behaviour, distinct from that predicted by classical polymer models⁴. To explain this anomalous stiffening, we introduce a hierarchical worm-like chain model that describes the chromosome as a heterogeneous assembly of nonlinear worm-like chains. Moreover, through inducible degradation of TOP2A⁵ specifically in mitosis, we provide evidence that TOP2A has a role in the preservation of chromosome compaction. The methods described here open the door to a wide array of investigations into the structure and dynamics of both normal and disease-associated chromosomes.

Main

The structure of eukaryotic chromosomes changes markedly as cells traverse the cell division cycle. In interphase, nuclear DNA has a diffuse appearance and individual chromosomes are not discernible. As cells enter mitosis, the replicated chromosomes condense into compact, cylindrical structures comprising two sister-chromatid arms that mature into the iconic

chromosome X shape in metaphase, in which the sister chromatids are held together only at the centromere. The sisters are then segregated to the nascent daughter cells in anaphase and telophase using force applied by the mitotic spindle^{6,7}. The prevailing model of chromosome organization posits that consecutive loops of chromatin are organized in a helical staircase conformation⁸ that is anchored to a central protein scaffold, with condensins I and II and TOP2A being key factors in mitotic chromosome formation^{3,9,10,11}. Although condensins have also been shown to be essential for the maintenance of a compacted chromosome structure^{2,12}, there are conflicting views on the role of TOP2A in this process^{5,13}.

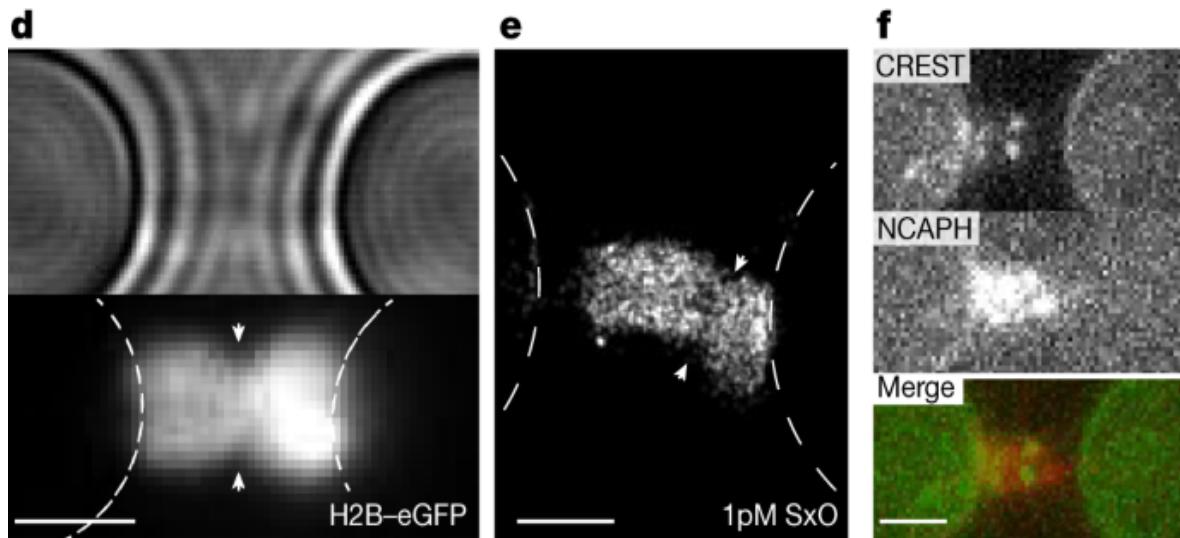
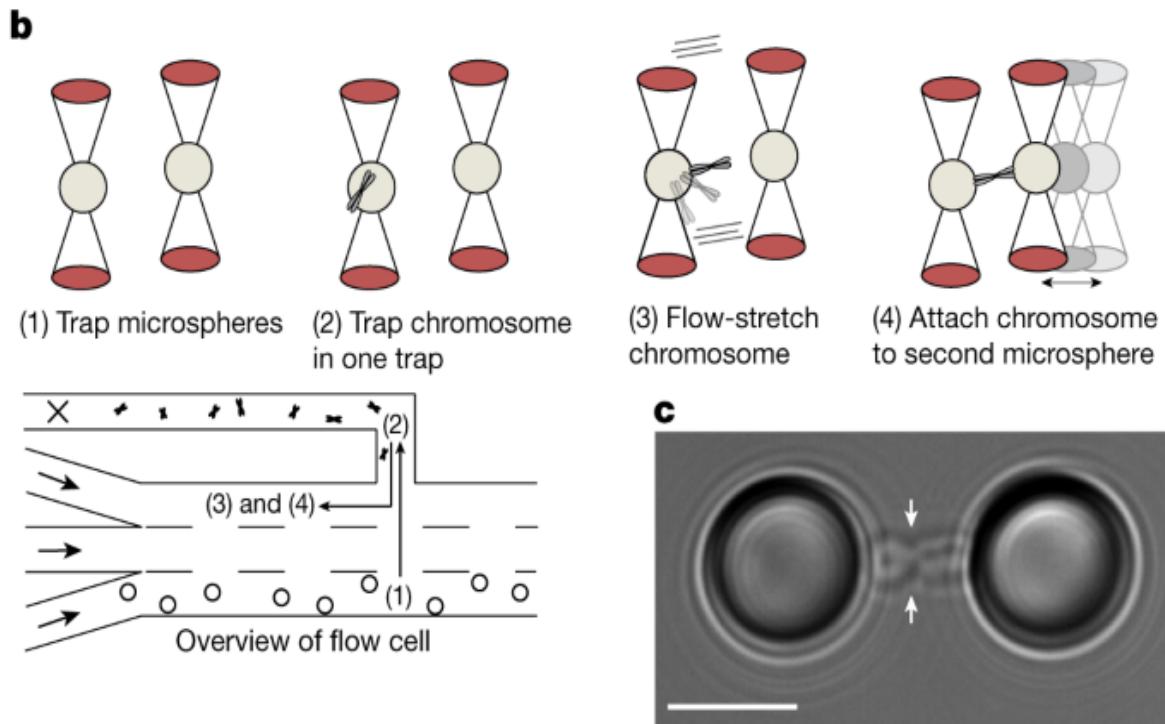
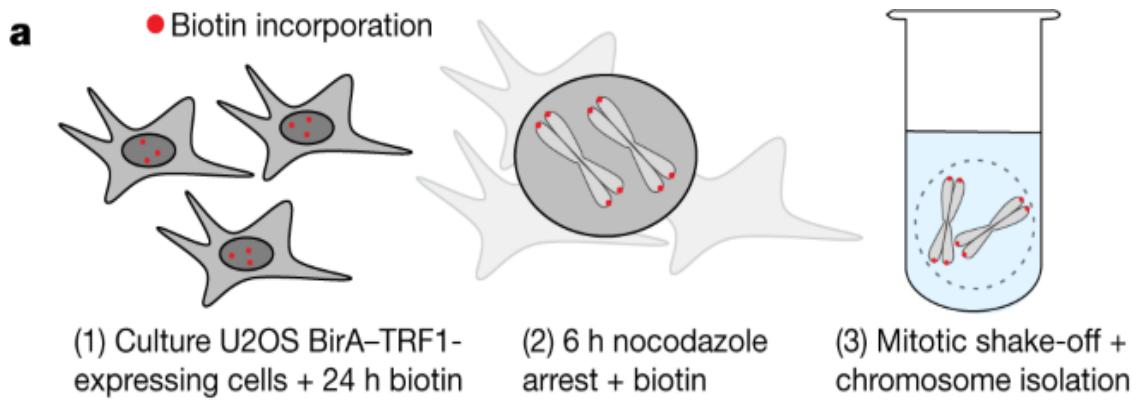
The mechanics and dynamics of many biomolecules have been elucidated by the use of micromechanical measurements, such as atomic force microscopy, and magnetic and optical tweezers¹⁴. Nevertheless, few mechanical studies have been performed on chromosomes¹⁵. A series of studies quantifying the mechanical stability of amphibian and human chromosomes, analysed by stretching chromosomes with micropipettes^{16,17,18}, revealed that they can be reversibly stretched by up to five times their native length by applying forces in the nanonewton range, and that depletion of condensins results in decreased chromosome stiffness¹⁹. Moreover, studies using Hi-C and super-resolution fluorescence microscopy^{20,21} have provided important information on chromosome structure and organization. To gain direct access to the dynamic structural features of chromosomes, while avoiding fixation and ensemble averaging, we introduce here a workflow to analyse the mechanics and architecture of mitotic chromosomes using a combination of optical tweezers, fluorescence microscopy and microfluidics, which readily allows the manipulation of individual native chromosomes with nanometre precision and piconewton force resolution.

Handling and visualizing chromosomes

To study native metaphase chromosomes using optical tweezers, we purified chromosomes with biotinylated telomeric ends, which served as molecular ‘handles’ for site-specific attachment to streptavidin-coated microspheres (Fig. 1a). Telomere-specific biotinylation was achieved through fusion of BirA protein to telomere repeat-binding factor 1 (TRF1)²². Biotin treatment

of either U2OS-BirA-TRF1 cells or HCT116 cells transduced with a TRF1-BirA lentivirus resulted in biotin incorporation at approximately 98% of telomeres ($n_{\text{tot}} = 1,434$; Extended Data Fig. 1a). We then optimized a chromosome isolation protocol²³ that yielded highly concentrated (around 10^6 chromosomes per ml) native mitotic chromosomes devoid of cell debris and cytoskeletal contaminants (see Methods), which were suitable for telomeric attachment between two microspheres held in optical traps (Fig. 1b, c, Extended Data Fig. 1b–d, Supplementary Video 1, Methods). This permitted us to handle and image chromosomes in a precisely controlled environment and accurately measure forces applied to the chromosome^{24,25}. We visualized the trapped chromosomes using either wide-field fluorescence imaging of eGFP-labelled histone H2B, allowing the individual sister chromatids and the centromeres to be discerned (Fig. 1d), or super-resolution imaging using binding-activated localization microscopy (BALM)²⁶, which permitted the visualization of individual fluorescent intercalators intermittently binding to a chromosome (Fig. 1e, Extended Data Fig. 1e). Our system is also compatible with multi-colour immunofluorescence analysis; for example, staining for histone H3 showed the expected pan-chromosome localization (Extended Data Fig. 1f), and staining for CREST and NCAPH revealed centromeres and chromosome scaffolds, respectively (Fig. 1f, Extended Data Fig. 1g, h, Methods). These different imaging strategies allowed us to confirm known features of chromosome organization.

Fig. 1: Workflow for chromosome attachment and visualization.



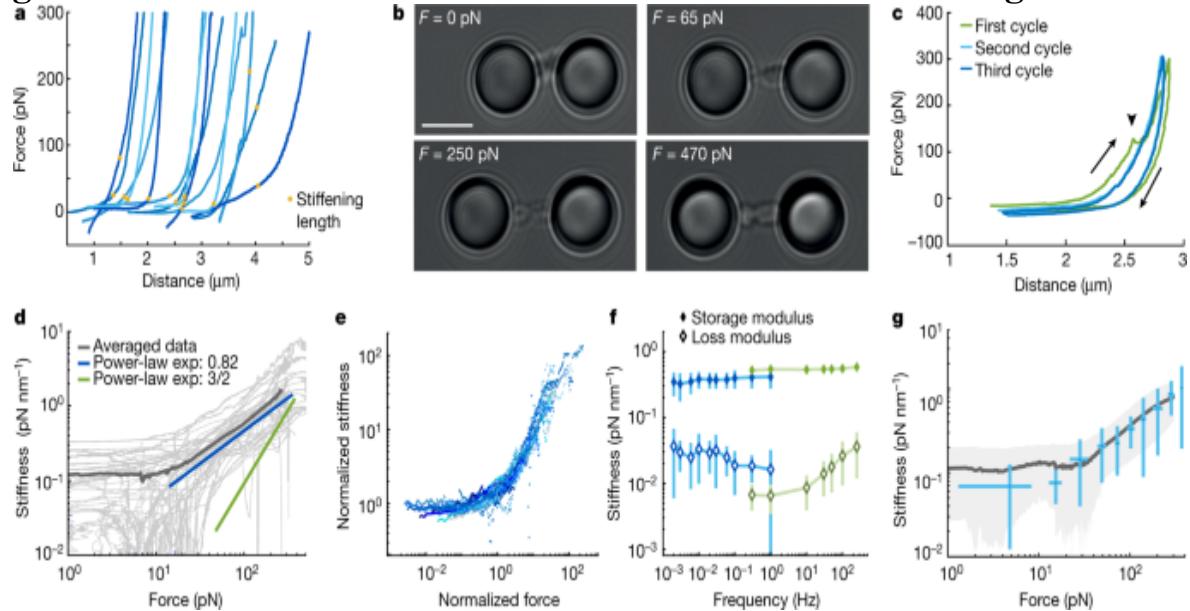
a, Diagram depicting the experimental workflow. After addition of biotin to U2OS cells expressing BirA–TRF1, proteins located at telomeric ends are covalently biotinylated. Chromosomes were then purified from cells arrested in prometaphase by treatment with nocodazole. **b**, Schematic depiction of chromosome attachment to microspheres in a microfluidic flow cell with parallel channels (bottom left). After trapping two streptavidin-coated microspheres (1), a chromosome is attached to one microsphere by exploiting the attraction force that the optical trap exerts on the chromosome (2). The chromosome is then flow-stretched (3), bringing it into the imaging plane, and attached to the second microsphere (4). **c**, Representative bright-field image of a mitotic chromosome showing telomeric attachment of the four chromatid ends. The centromeric region is discernible as a constriction (arrows). Scale bar, 4 μm . **d**, Representative fluorescence image of H2B–eGFP (bottom) and corresponding bright-field image (top). Note that the chromosome in **d** was positioned to be in the focal plane for fluorescence imaging and not bright-field imaging. Scale bar, 2 μm . **e**, Representative BALM super-resolved image of SYTOX orange (SxO) binding events. The centromeric region is indicated with arrowheads. Scale bar, 1 μm . **f**, An example of immuno-staining of NCAPH and CREST to show the localization of condensin I along the chromatid scaffold and two foci that reveal the position of the centromeres, respectively. Scale bar, 2 μm .

Mechanical properties of chromosomes

Optical manipulation enables mechanical features of chromosomes to be characterized with very high resolution. To achieve this, we first recorded force-extension curves by separating the two optical traps at a constant velocity of less than 0.2 $\mu\text{m s}^{-1}$ (Fig. 2a, b, Supplementary Videos 2, 3). For forces up to typically 10–50 pN, the force-extension behaviour of individual chromosomes was approximately linear, albeit with a large variability in stiffness between different chromosomes. By contrast, at higher forces, the chromosomes exhibited pronounced nonlinear stiffening, such that the force increased markedly with increasing chromosome extension. We determined chromosome length at the onset of this stiffening (stiffening length), which showed a broad distribution that is likely to reflect the known variability in the size of human chromosomes ($2.5 \pm 1.0 \mu\text{m}$ and $2.8 \pm 1.7 \mu\text{m}$ for U2OS and HCT116 chromosomes, respectively, mean \pm s.e.m.; Extended Data Fig.

[2a](#)). Moreover, the force-extension response was reversible at forces up to 300 pN (Fig. [2c](#), Extended Data Fig. [2b](#)), as reported in micropipette aspiration studies^{[18,19](#)}.

Fig. 2: Mechanical characterization of chromosome stretching.



a, Representative series of force-extension curves depicting a linear stiffness regime for low forces ($F < 10 \text{ pN}$) and a nonlinear stiffness regime at higher forces ($F > 20 \text{ pN}$), in which the stiffness of the chromosomes increases with force. Stiffening lengths are indicated with dots. **b**, Representative bright-field images of stretching a U2OS chromosome. Scale bar, $4 \mu\text{m}$. **c**, Three consecutive cycles of elongation and retraction (direction depicted by arrows). Abrupt declines in the force response (arrowhead) are suggestive of small rupture events and are most abundant during the first stretch cycle (23% of curves; $n = 155$ compared to 14% of third stretches; $n = 79$). **d**, Individual differential stiffness curves (light grey; $n = 44$) and average curve (dark grey) show a linear stiffness regime up to around 10 pN followed by power-law scaling in the regime between 20 and 200 pN, with a scaling exponent (exp) of $\gamma = 0.82 \pm 0.05$ (blue). The green line depicts scaling of 3/2 as expected for a WLC. **e**, Datasets show a universal form of the nonlinear stiffening after rescaling by K_0 and F_c , independent of the differences among chromosomes. **f**, The storage modulus (filled diamonds) and loss modulus (open diamonds) determined at a force of 50 pN with two methods of distance detection for the lower (blue; $n = 8$) and the higher

(green; $n = 11$) frequencies ([Methods](#)). **g**, Oscillation data measured at 1 Hz ($n = 14$), overlapped with the differential stiffness derived from force-extension experiments (dark grey line, mean; light-grey shaded region, s.d.). In **f**, **g**, data represent mean values \pm s.d. The data in **a**, **d**, **e**, **g** are from the third extension cycle.

To quantify the stiffening of chromosomes at high force, we determined the differential stiffness K by evaluating the numerical derivative of the force F with respect to the extension d (Extended Data Fig. [2c](#), [Methods](#)). Beyond a critical force F_c the differential stiffness increased following a power-law dependency; $K \sim F^\gamma$ (Fig. [2d](#)), in which the stiffening exponent γ characterizes how sensitive the stiffening is to force. Classical models for polymers predict a power-law stiffening, with the freely jointed chain (FJC) being characterized by the stiffening exponent $\gamma = 2$ and the worm-like-chain (WLC) by $\gamma = 3/2$ (refs. [4,27,28,29](#)). However, we observed a markedly weaker stiffening exponent compared to these classical models (Fig. [2d](#)), with $\gamma = 0.82 \pm 0.05$ for U2OS chromosomes. Despite a variability in initial stiffness between different chromosomes, all stiffness-force curves could be approximately collapsed onto a single master curve by scaling them to initial stiffness and critical force F_c (Fig. [2e](#), [Methods](#)), which showed that the mechanical behaviour amongst chromosomes is consistent. To investigate whether this anomalous behaviour reflects structural remodelling under load—for example, owing to dynamic cross-linking such as that described for F-actin networks^{[30,31,32](#)}—we measured the frequency dependence of the stiffening response. Using microrheology measurements, we quantified this viscoelastic response by applying a fixed pre-tension F to chromosomes, and then a small-amplitude distance oscillation with frequency ω (refs. [28,32](#)). The differential force response is captured by the storage modulus $K'(F,\omega)$ and the loss modulus $K''(F,\omega)$, which characterize the elastically stored and the viscously dissipated mechanical energy, respectively ([Methods](#)). We determined that the storage modulus was constant over five orders of magnitude in frequency and was 10- to 100-fold larger than the loss modulus (Fig. [2f](#)). When measurements were performed at varying pre-tensions, the storage modulus was consistent with the differential stiffness derived from the force-extension data (Fig. [2g](#)). These results show that chromosomes have a predominantly elastic mechanical response over a wide range of frequencies, which is inconsistent with substantial dynamic remodelling

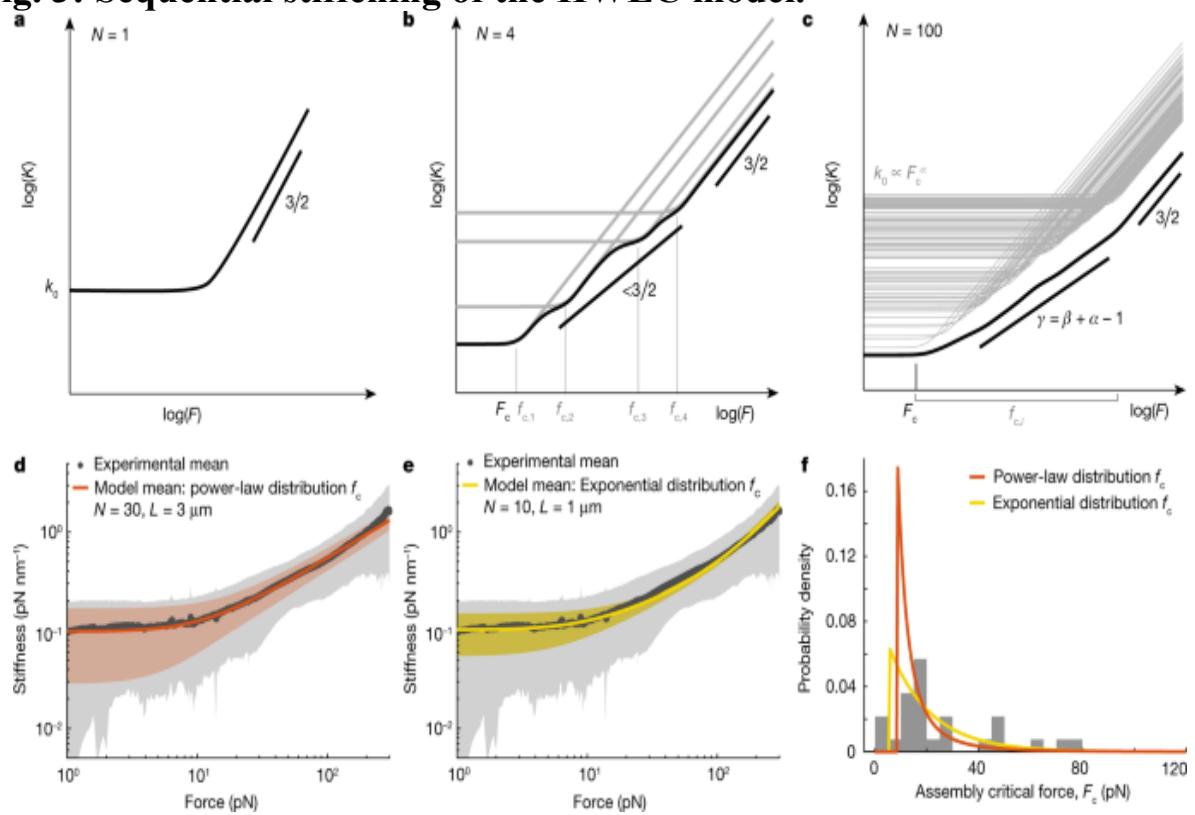
occurring on these time scales. Thus, we attribute the notably weak power-law stiffening response to an intrinsic, nonlinear elastic response of the chromosome.

Chromosome hierarchical mechanics

Models of homogeneous polymers with finite extensibility d_c , such as the FJC or WLC, exhibit a nonlinear response at high force, such that $F \sim (d - d_c)^{-\delta}$, with $\delta > 0$ (refs. 4,27). This divergent force-extension behaviour implies $K \sim F^\gamma$ with a stiffening exponent $\gamma = (\delta + 1)/\delta$, which is strictly larger than 1. Thus, the weak anomalous stiffening exponent ($\gamma < 1$) that we determined suggests that the nonlinear mechanical response lies in the inherent heterogeneous nature of chromosomes. Heterogeneity of chromosomes is known to arise owing to the presence of specialized proteinaceous structures, such as centromeres, as well as the inherent differences between euchromatic and heterochromatic regions, which is reflected in the appearance of G-banding patterns after Giemsa staining^{33,34}. To capture this heterogeneous nature of chromosomes, we propose a hierarchical worm-like chain (HWLC) model: an assembly with different structural elements represented by a series of WLCs with distinct contour and persistence lengths, which leads to an emergent nonlinear behaviour that is different from the response of the individual elements. Upon mechanical loading, these elements respond in a force-dependent hierarchy that leads to sequential stiffening (Fig. 3a–c, Extended Data Fig. 3). At low force, the response is dominated by the softest element. At higher forces, the stiffness of the next softest element dominates, and so on. The hierarchical nature of this model is characterized by the distribution $P(f_c)$ of internal critical forces of the individual elements. We compared our data to two classes of HWLC models: in one, we randomly drew the critical forces of the components from a power-law distribution $P(f_c) \propto f_c^{-\beta}$ (Fig. 3d); and, in the other, we drew them from an exponential distribution \\(P\\left(\\{ f \\} _{\\{ \\rm{c} \\}} \\right) \\propto \\{ \\rm{e} \\} ^{-\\{ f \\} _{\\{ \\rm{c} \\}} / \\{ f \\} _{\\{ \\rm{c} \\}} ^{*}}) (Fig. 3e, Supplementary Note 1, Extended Data Fig. 4). Although both models showed anomalous stiffening, the power-law distribution offered a closer agreement to the spread in chromosome stiffness (Fig. 3d,e, Extended Data Fig. 5). The distribution of

critical forces of the assembly for both models is in accord with experimental observations (Fig. 3f, [Methods](#)). However, only a power-law distribution for critical forces (Fig. 3d) provides a genuine power-law stiffening of the HWLC with an exponent, $\gamma = \beta + \alpha - 1$, in which α is set by the relation between the initial stiffness and the critical force of each element, $k_0 \propto f_c^\alpha$ ([Supplementary Note 1](#)). We conclude that the HWLC model can quantitatively account for the observed nonlinear mechanical response of human chromosomes.

Fig. 3: Sequential stiffening of the HWLC model.



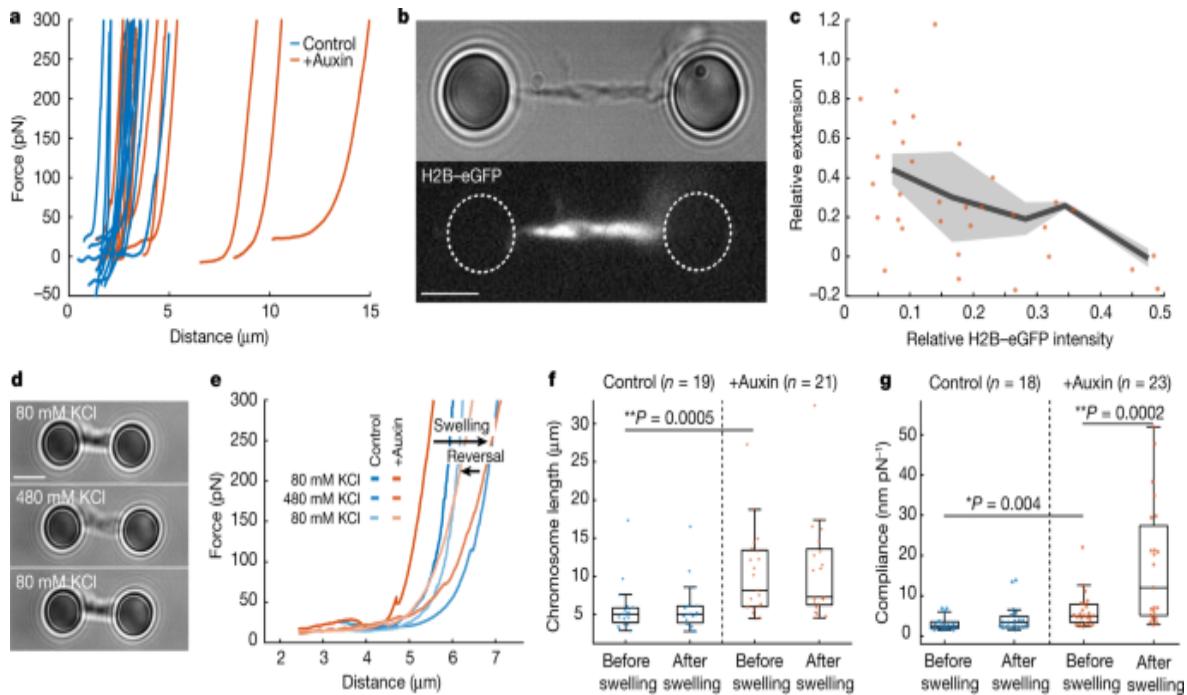
a, Stiffness-force curves of a single WLC transition from a constant stiffness directly towards a strain stiffening regime with a power-law scaling of 3/2. **b**, Stiffness-force curves of an assembly of four WLCs (individual WLC response in grey) in series result in an overall response (black) with an irregular transition zone to nonlinear stiffening that has a weaker stiffening behaviour than an individual WLC. The critical forces of the elements $f_{c,i}$ and of the assembly F_c are indicated on the x axis. **c**, Stiffness-force curve of an assembly (black) of 100 serial WLCs (individual WLC response in grey)

with critical forces drawn from a power-law distribution, $P(f_c) \propto f_c^{-\beta}$, resulting in a smooth power-law transition regime with an exponent of $\gamma = \beta + \alpha - 1$. **d, e**, Comparison of the experimental mean and s.d. of U2OS chromosomes (grey) with the mean and s.d. of the HWLC model, with $P(f_c) \propto f_c^{-\beta}$ with $\beta = 0.86$ ($N = 30$ and $L = 3 \mu\text{m}$) (orange) (**d**) and $(P(\{f\}_{\{\{\text{rm}\{c\}\}\}}) \propto \exp\{-\{f\}_{\{\{\text{rm}\{c\}\}\}}/\{f\}_{\{\{\text{rm}\{c\}\}\}}^{*\beta}\})$ with $\{f\}_{\{\{\text{rm}\{c\}\}\}}^{*\beta} = 160 \text{ pN}$ ($N = 10$ and $L = 1 \mu\text{m}$) (yellow) (**e**). **f**, The theoretical distribution of assembly critical forces F_c for the model parameters from **d** (orange) and **e** (yellow), superimposed with experimentally determined assembly critical forces of U2OS chromosomes ($n = 29$) ([Methods](#)).

Mechanical role of TOP2A

To investigate the relationship between chromosome structure and mechanics, we depleted TOP2A, which is required for mitotic chromosome formation^{3,11,35,36}, and might also have a specific structural role^{5,10,11,37}. We analysed chromosomes from human HCT116 CDK1as cells, in which TOP2A could be depleted during prometaphase using a ‘degron’ system⁵ (Extended Data Fig. [6](#), [Methods](#)). We observed two populations of TOP2A-depleted chromosomes—one with a stiffening length equivalent to that of TOP2A-containing chromosomes, and another elongated and hypo-condensed population (Fig. [4a,b](#), Extended Data Fig. [7a](#)). Hypo-condensed chromosomes were present at a higher frequency in chromosome spreads than in the tweezers (Extended Data Fig. [6f](#), Supplementary Note [2](#)). Notably, depletion of TOP2A only led to minor quantitative changes in stiffening behaviour, which still followed the HWLC model (Extended Data Fig. [7b,c](#), [Methods](#)). Fluorescence images of hypo-condensed, H2B-eGFP-labelled, TOP2A-depleted chromosomes showed a heterogeneous distribution of histones along the chromosome arms (Fig. [4b](#)). After stretching, the brighter, chromatin-dense regions extended less than the less dense regions (Fig. [4c](#), Extended Data Fig. [7d](#)), a heterogeneity consistent with our HWLC model.

Fig. 4: Mechanical properties of TOP2A-depleted chromosomes.



a, Stretching curves of control chromosomes (blue) and TOP2A-depleted chromosomes (orange). **b**, Representative bright-field image of a TOP2A-depleted chromosome and corresponding H2B immunofluorescence image. Scale bar, 4 μm . **c**, The relative extension of darker and brighter H2B–eGFP regions as a function of relative intensity (grey line and shaded region: mean \pm s.e.m.). **d**, Representative bright-field images of a chromosome in its original buffer containing 80 mM KCl and subsequent images after flushing in buffer containing 480 mM KCl followed by buffer containing 80 mM KCl. **e**, Force-extension curves of a control and a TOP2A-depleted HCT116 chromosome before, during and after exposure to high-salt buffer (480 mM KCl). Changes in length (**f**) and compliance (**g**) of chromosomes before and after the exposing control and TOP2A-depleted chromosomes to high-salt buffer. Two-sided Wilcoxon rank-sum test, * $P < 0.05$; ** $P < 0.01$. Centre, median; box, 25th to 75th percentile; whiskers, minimum and maximum data points (not considered as outliers).

To investigate whether TOP2A has a role in preserving the structure of condensed chromosomes, we induced chromosome decompaction (swelling) and re-compaction by alternating the KCl concentration^{2,38} (Fig. 4d, Supplementary Video 4). During decompaction, we observed a strong elongation and softening of the chromosomes (Fig. 4e). After re-compaction, both control and TOP2A-depleted chromosomes returned to their initial

length (Fig. 4f). Nevertheless, there was a significant increase in the compliance (inverse stiffness) of the TOP2A-depleted chromosomes after re-compaction (from $6 \pm 1 \text{ nm pN}^{-1}$ to $15 \pm 3 \text{ nm pN}^{-1}$ at 200 pN; $P = 0.0002$), unlike in control chromosomes (Fig. 4g, Methods). Similarly, we observed no change in the stiffening exponent of control chromosomes, whereas that of the hypo-condensed chromosomes (stiffening length greater than 5 μm) decreased significantly from 1.1 ± 0.2 to 0.4 ± 0.1 ($P = 0.0017$) (Extended Data Fig. 7e). Within the HWLC framework, such a decrease in the stiffening exponent is interpreted as a structural perturbation with a flatter distribution $P(f_c)$ of the critical forces of the components. Moreover, the force-extension curves of chromosomes after decompaction did not change after prolonged exposure to high salt, indicating that there was no loss of material on the timescale of the experiment (Extended Data Fig. 7f). We conclude that TOP2A assists in the restoration of the chromosome to its original structure after perturbation. Thus, TOP2A is not only indispensable for chromosome condensation, but it is also important for the preservation of mitotic chromosome structure.

Conclusions

Here, we have introduced a strategy to study mitotic chromosomes using optical tweezers. We have successfully visualized individual human chromosomes at high resolution and analysed their mechanical parameters with a very high control of applied force. The chromosomes exhibited a nonlinear stiffening response with a power-law exponent considerably lower than that predicted by established polymer models. Although the initial stiffness and length of chromosomes was variable, the anomalous nonlinear stiffening was robust, suggesting that this is an inherent characteristic of chromosomes. This stiffening behaviour is distinct from the linear force-extension that has been reported previously using micropipettes^{18,19}, which probably stems from differences in force resolution and chromosome attachment. To explain the observed nonlinear mechanics, we developed a HWLC model based on sequential stiffening of hierarchical elements within a heterogeneous chromosome (Extended Data Fig. 3). Hence, the anomalous stiffening behaviour of a chromosome emerges from its intrinsic heterogeneity. It is tempting to speculate that such a hierarchy of nonlinear mechanical elements could be beneficial for maintaining the structural

integrity of chromosomes by limiting the deformation of individual elements.

A future challenge is to identify the molecular basis for this HWLC model, and establish whether it could relate to patterns of heterochromatin and euchromatin³⁹, A-T and G-C content or the distribution of structural proteins along the chromosome^{33,40,41}. A notable feature after depletion of TOP2A—the most abundant non-histone protein in metaphase chromosomes—is a reproducible shift in the stiffening response after perturbing chromosome structure. This is in stark contrast to other examples of elastic stress-stiffening polymer assemblies, in which non-destructive structural perturbations do not alter the stiffening exponent. Of note, the HWLC model can capture such a change in the stiffening exponent in terms of a structural modification. It has been hypothesized that TOP2A has a structural role by stabilizing chromatin loops through simultaneously binding two DNA duplexes in a closed gate⁵. Indeed, we find that TOP2A is essential to sustain chromosome mechanics after salt-induced chromosome expansion. The ability to manipulate and image chromosomes under controlled conditions makes our method suitable for investigating the structural and mechanical roles of other chromosomal proteins. We foresee that, by selective depletion or exposure to inhibitors, more insight will be obtained into how proteins compact and maintain mitotic chromosomes. In summary, we have transformed an already-powerful single-molecule technique into a quantitative and versatile method for investigating the mitotic chromosome.

Methods

Cell lines and cell culture

All cell lines were cultured in DMEM supplemented with 10% fetal bovine serum (FBS) and penicillin–streptomycin in a humidified incubator at 37 °C and 5% CO₂. Unless indicated otherwise, all cell lines were obtained from and authenticated by the ATCC by karyotyping and STR profiling. The U2OS TRF1-BirA cell line²² was a gift from R. J. O’ Sullivan and was authenticated by karyotyping. Endogenous H2B in the U2Os TRF1-BirA cell line was tagged with eGFP, as described previously⁵. To rapidly deplete

TOP2A, we used a HCT116 TOP2A-mAID cell line that also expressed H2B-eGFP, facilitating chromosome identification. The HCT116 TOP2A-mAID H2B-eGFP cell line was a gift from D. F. Hudson and authentic by karyotyping, and was described in a previous report⁵. All cell lines were routinely tested for mycoplasma and shown to be negative. To achieve tight temporal control over cell synchrony, the HCT116 TOP2A-mAID H2B-eGFP cell line was modified for CDK1as chemical genetics by knock-in of CDK1as and knockout of endogenous CDK1, as described previously⁴² and based on another previous report⁴³. The constructs for CDK1as were gifts from W. Earnshaw (Addgene 118596 and 118597) and from Z. Izsvák (Addgene 34879). We determined that treatment with 0.25 μM 1NM-PP1 (529581, Sigma-Aldrich) for 16 h efficiently arrests HCT116 TOP2A-mAID H2B-eGFP CDK1as cells at the G2–M boundary (Extended Data Fig. [6a, b](#)). The incubation time and concentration of 1NM-PP1 were optimized by propidium iodide flow cytometry (Extended Data Fig. [5a](#)), performed as described before⁵. Efficient release from the arrest was achieved with two wash cycles by centrifugation with preheated medium. A Neon transfection system (Thermo Fisher Scientific) was used for transfections of HCT116 and U2OS cell lines according to the manufacturer’s recommendations. 1NM-PP1 and nocodazole were purchased from Sigma-Aldrich. The synthetic auxin indole-3-acetic acid (IAA) sodium salt (sc-215171, Santa Cruz) was used. Six days before chromosome isolation, HCT116 TOP2A-mAID CDK1as cells were transduced with lentiviruses introducing TRF1-BirA into the genome. These cells were treated for 16 h with 0.25 μM 1NM-PP1, before release into 100 ng ml⁻¹ nocodazole (Sigma-Aldrich) with or without 500 μM auxin for 4 h, to arrest cells in prometaphase and deplete TOP2A, respectively. Mitotic cells were detached by shaking and chromosomes were isolated from this population (Extended Data Fig. [6c](#)). Chromosome spreads were performed as described previously⁵ and showed an altered chromosome morphology following exposure to auxin in accordance with what was reported⁵. Approximately 75% of TOP2A depleted chromosomes appeared hypocondensed compared to 5% of control chromosomes (Extended Data Fig. [6d](#)). Immunostaining of TOP2A on chromosome spreads was not detectable in auxin-treated samples, confirming efficient depletion of TOP2A (Extended Data Fig. [5e](#)).

Lentiviral production and transduction

Third-generation lentiviral particles were generated for integration of BirA-TRF1. HEK293T cells were grown with 25 µM chloroquine diphosphate (Sigma-Aldrich) for 5 h before being transfected with plasmids pMD2.G, pMDLg/pRRE and pRSV-Rev (Addgene 12259, 12251 and 12253, deposited by D. Trono⁴⁴) and a transfer plasmid for BirA-TRF1 integration. A Calphos mammalian transfection kit (Clontech) was used for transfections according to the manufacturer's protocol. Eighteen hours after transfection, the medium was replaced with fresh medium. Forty-eight hours after transfection, the growth medium was collected and centrifuged at 500g for 5 min, and the supernatant containing viral particles was filtered through a 0.45-µm membrane before being concentrated 10× using an Amicon Ultra-15 100 kDa centrifugal unit (Merck-Millipore). The viral concentrate was snap-frozen and stored at -80 °C. For lentiviral transduction, a T-75 flask of 75% confluent HCT116 TOP2A-mAID CDK1as cells was incubated with 7.5 µg ml⁻¹ polybrene in 3 ml 10× lentiviral concentrate and 7 ml growth medium for 1 h with mixing every 15 min. Cells were then seeded in a T-175 flask and the culture was expanded before chromosome isolation.

Chromosome isolation

A previously reported method, with modifications, was used to isolate mitotic chromosomes in large quantities with minimal contamination with cell debris²³. In brief, cells were grown with 12.2 mg l⁻¹ biotin (Sigma-Aldrich) for 24 h before isolation. On the day of isolation, 8–10 T175 flasks of cells were treated for 4 h with 200 ng ml⁻¹ nocodazole (Sigma-Aldrich) and then mitotic shake-off was used to enrich for mitotic cells, resulting in 1 × 10⁷–2 × 10⁷ mitotic cells. The mitotic cells were centrifuged at 300g for 5 min, resuspended in 10 ml 75 mM KCl and 5 mM Tris-HCl (pH 8.0) and then incubated for 10 min at room temperature. All subsequent steps were carried out at 4 °C. Cells were centrifuged at 300g for 5 min and then resuspended in 8 ml polyamine (PA) buffer (15 mM Tris-HCl (pH 8.0), 2 mM EDTA, 0.5 mM EGTA, 80 mM KCl, 20 mM NaCl, 0.5 mM spermidine, 0.2 mM spermine and 0.2% Tween-20) for U2OS cells and a PA* buffer (15 mM Tris-HCl (pH 7.4), 0.5 mM EDTA-K, 80 mM KCl, 1 mM spermidine, 0.4 mM spermine and 0.1% Tween-20) for HCT116 cells, both supplemented with Complete mini protease and PhosSTOP phosphatase inhibitor cocktails (Roche). This suspension was then lysed in a Dounce

homogenizer using 25 strokes with a tight pestle. The suspension was cleared twice of cell debris by centrifugation at 300g for 5 min.

Chromosomes were purified using a glycerol step gradient containing two layers (60% and 30% glycerol in PA). After centrifugation at 1,750g for 30 min, the chromosomes were collected from the 60% glycerol fraction and stored at -20 °C in around 60% glycerol in PA buffer at a concentration of 10⁶–10⁷ chromosomes per ml. Chromosomes could be stored for up to two months without undergoing any noticeable change in mechanical properties.

Dual trap optical tweezers with wide-field fluorescence

The dual trap optical set-up was described previously²⁴. In brief, two optical traps were created using a 20 W, 1064 nm CW fibre laser (YLR-20-LP-IPG, IPG Photonics). Two traps were created by splitting the laser beam into two paths using a polarizing beam splitter cube and could be steered independently using one accurate piezo mirror (Nano-MTA2X10, Mad City Labs) and one coarse positioning piezo step mirror (AG-M100N). After the two paths were recombined, they were coupled into a Nikon microscope body using two 300 mm lenses, and focused in the flow cell with a 1.2 NA water immersion objective (Nikon, Plan apo VC NA1.2). Back-focal plane interferometry was used to measure forces, and bead tracking was performed by LED illuminated bright-field imaging on a CMOS camera (DCC1545M, Thorlabs). Wide-field epifluorescence was achieved by illumination with 488, 532, 561 and 639 nm lasers (Cobolt 06-01 Series) and detection by separation of the emission light using an OptoSplit III (Cairn Research) and imaging on an EMCCD camera (iXon 897 Life, Andor Oxford Instruments Technology).

Microfluidics and flow cell preparation

A microfluidic flow-system (u-Flux, LUMICKS B.V.) was used to insert solutions into a five-channel flow cell (LUMICKS B.V.; Fig. [1b](#)). Before each experiment, bleach cleaning was performed to remove residual debris from flow cell, followed by sodium thiosulfate neutralization. Passivation was performed to reduce chromosome attachment to tubing and flow cell walls by incubation for 1 h with 0.05% casein solution, followed by excessive rinsing with PA buffer. Chromosomes diluted in PA buffer (10–20

μl in $500\mu\text{l}$) were inserted into a side channel of the flow cell (Fig. [1b](#)). Streptavidin-coated polystyrene microspheres (diameter: $4.6\mu\text{m}$, Spherotech) in PA buffer ($4\mu\text{l}$ in $300\mu\text{l}$) were inserted in one of the main channels. Other channels were filled with PA buffer unless stated otherwise.

Chromosome attachment and force-extension

To facilitate attachment of the biotinylated chromosome between two streptavidin-coated microspheres (diameter: $4.6\mu\text{m}$), one trapped microsphere was brought into the proximity of a chromosome in solution, resulting in attachment of the telomeric end of the chromosome to the microsphere (Fig. [1b](#), Extended Data Fig. [1b](#)). Next, the microspheres were moved to another microfluidic channel and fluid flow was activated (Fig. [1b](#)). The chromosome attached to one of the microspheres was flow-stretched to confirm correct attachment (Extended Data Fig. [1b,c](#)) and then brought into the proximity of the other microsphere to induce attachment of the other chromosome end (Fig. [1c](#)). Note that owing to the relatively small cross-section of chromosomes compared to the microspheres, both telomeric ends from one sister chromatid would attach to the microsphere occasionally. Non-biotinylated chromosomes showed only very limited attachment to the microspheres (Extended Data Fig. [1d](#)).

Immunofluorescence

Chromosomes were incubated overnight at 4°C with primary antibody in a concentration of $5\mu\text{g ml}^{-1}$ and were subsequently diluted fivefold in PA buffer and stored for 1 h at 4°C . Next, chromosomes were incubated with secondary antibody in a concentration of $5\mu\text{g ml}^{-1}$ for 1 h at room temperature. After addition of PA buffer to dilute the sample again by fivefold, chromosomes were stored for 30 min at 4°C . To remove excess antibody, chromosomes were centrifuged at 750g for 5 min on a $20\mu\text{l}$ glycerol cushion. The supernatant was then removed, leaving around $100\mu\text{l}$ chromosome solution that could be used for imaging. Primary antibodies were anti-NCAPH (1:100, HPA002647, Sigma Aldrich), CREST anti-sera (1:200 HCT-0100, Immunovision), anti-TRF2 (1:100, sc-9143, Santa Cruz), anti-H3S10 (1:400, 06-570, Sigma-Aldrich) and anti-H3-Alexa Fluor 647 (1:200, 15930862, Thermo Fisher Scientific). Secondary antibodies were

anti-rabbit IgG-Alexa Fluor 647 (1:500, A-21244, Thermo Fisher Scientific), anti-rabbit IgG-Alexa Fluor 568 (1:500, A-11011, Thermo Fisher Scientific) and anti-human IgG-Alexa Fluor 488 (1:500, A-11013, Thermo Fisher Scientific). Biotinylated TRF1 was detected using streptavidin–Alexa Fluor 568 (1:200, S11226, Invitrogen).

Immunoblotting

SDS-PAGE and immunoblotting was performed as described previously⁵. In brief, cell pellets were lysed in RIPA buffer containing cOmplete Mini EDTA free (Roche) on ice for 20 min. Samples were then sonicated in a water-cooled Bioruptor Pico (Diagenode) and centrifuged at 21,000g for 15 min at 4 °C. Protein concentration was determined using a Pierce BCA protein assay kit (Thermo Fisher Scientific). Forty micrograms of protein was loaded per well. The primary antibodies were anti-CDK1 (1:1,000, ab133327, Abcam), anti-Myc (1:1,000, sc-40, Santa Cruz) and anti-histone H3.3 (1:5,000, ab176840, Abcam). The secondary antibodies were anti-mouse IgG peroxidase conjugate (1:10,000, A4416, Sigma-Aldrich) and anti-rabbit IgG peroxidase conjugate (1:10,000, A6154, Sigma-Aldrich).

Determination of differential stiffness, stiffening length and compliance

To calculate the differential stiffness from force-distance curves, the force distance curve was first smoothed using a moving average with a window size of 1/15 of the total data points in the force curve, followed by numerical differentiation of force with respect to distance. To determine the onset of stiffening, the plateau stiffness was determined as the most likely stiffness at forces below 50 pN, as estimated from the maximum of a kernel density estimate of the stiffness values. The onset of stiffening was then determined as the point at which the stiffness exceeds the plateau stiffness by one standard deviation of all stiffnesses below 50 pN. To determine the compliance at 200 pN, the inverse of the stiffness of the chromosome at a force of 200 pN was used.

Collapse of stiffness-force curves

To achieve a collapse of the stiffness-force curves they were normalized on a log-log-scale. Therefore, curves were interpolated to a logarithmic force scale to get evenly spaced data after taking the logarithm. In addition, negative values for force and stiffness were discarded. Then the logarithms of force $\ln(F)$ and stiffness $\ln(K)$ were calculated and fitted with a piecewise function $y = \ln(K_0)$ for $x \leq \ln(F_c)$ and $y = c - \ln(F_c) + \ln(K_0)$ for $x > \ln(F_c)$ to determine the initial stiffness K_0 and the critical force F_c . If the determined parameters for K_0 and F_c were in the range of the stiffness-force curve, the curves normalized by K_0 and F_c were plotted in a double-logarithmic plot to achieve the collapse. The criteria that K_0 and F_c had to be positive and within the range of the stiffness-force curve were met by 29 out of 44 curves.

Microrheology

Oscillations of the optical trap were generated by applying a sinusoidal voltage to the analogue input of the piezo mirror controller (Nano-Drive, MCL) to apply the oscillation on top of the digitally controlled mirror position. The signal was first generated digitally using Labview (National Instruments). The analogue voltage was then produced with a digital analogue converter (DAQ, National Instruments). Oscillations were produced with an amplitude of 25 mV corresponding to a trap displacement of roughly 200 nm. The frequency of the oscillation was varied between 2 mHz and 100 Hz. When the frequency was varied in the experiment, the pre-tension was kept constant at 50 pN (Fig. 2f). Experiments for different pre-tension were performed with a frequency of 0.1 Hz (Fig. 2g). To avoid limitations by the frame rate of the bead tracking camera at higher frequencies (>1 Hz), the bead position at high frequencies was calculated from the force and the trap position (Fig. 2f, green line). Data analysis of the oscillations was performed in MATLAB (Mathworks). First, the bead-bead distance and the force were synchronized on the basis of the position of the stationary bead where the force was measured, based on the cross-correlation between the bead position from bead tracking and the measured force. Then the oscillatory data were analysed following a previously described procedure⁴⁵. In brief, both the force and the bead-bead separation were detrended and fitted with a sine function with a fixed frequency set to

the experimental frequency and an additional offset. Then the complex stiffness was calculated as $\langle k = \frac{A_F}{\omega_F} \rangle = A_d \cdot e^{i(\varphi_F - \varphi_d)}$ with the amplitude and the phase of the force oscillation A_F and φ_F , and the amplitude and the phase of the distance oscillation A_d and φ_d , respectively.

Calculating HWLC force responses

Model curves in Fig. 3d,e were constructed by first defining a distribution for each system parameter. For simplicity, the number of sub-chains, N , and the length of each sub-chain, L/N , were kept constant. A power-law or exponential distribution with cut-offs was chosen for the sub-chain critical force, f_c . Given these distributions, we analysed the responses of 500 HWLC configurations, each constructed by sampling N values of f_c . The force-response of each configuration was computed by summing the extensions of each sub-chain at a given force, found by numerically solving the flexible WLC equation (Supplementary Note 1). The force-response curve was then numerically differentiated, and the mean and standard deviation of the stiffness-force curve were compared to experimental data.

Distribution of F_c

The distribution of the critical force of a HWLC assembly, F_c , corresponds to the force at which its softest element starts stiffening. Hence, for a given distribution, $P(f_c)$, F_c is distributed as the minimum of N independent samples. Its cumulative distribution function (CDF) satisfies $P(F_c \leq x) = 1 - \left(1 - P(f_c > x)\right)^N$. This expression can be differentiated to yield the probability density function (PDF), $P(F_c = x) = N \cdot P(f_c > x)^{N-1}$. Figure 3f shows $P(F_c = x)$ for the power-law distribution with cut-offs and for an exponential distribution with cut-offs. The experimental critical forces were determined as described above.

Statistics and reproducibility

Average values and errors were represented as mean \pm s.e.m. unless indicated otherwise. Differences in populations are tested using a two-sided Wilcoxon rank-sum test, where $P < 0.05$ is regarded as significant (*) and $P < 0.01$ as highly significant (**). The sample sizes for representative microscopy images are as follows: Fig. 1c n = 91, Fig. 1d–f n = 3, Fig. 2b n = 91, Fig. 4b n = 5, Fig. 4d n = 20, Extended Data Fig. 1a n = 24, Extended Data Fig. 1b, c n = 3, Extended Data Fig. 1e n = 10, Extended Data Fig. 1f–h n = 3, Extended Data Fig. 6a n = 2, Extended Data Fig. 6e (amount of cells) n = 117 (15 min), n = 120 (20 min), n = 98 (25 min), n = 121 (30 min), n = 110 (40 min), n = 115 (50 min), Extended Data Fig. 6g n = 16 (control), n = 19 (+auxin), Extended Data Fig. 7d n = 5.

Reporting summary

Further information on research design is available in the [Nature Research Reporting Summary](#) linked to this paper.

Data availability

The data supporting the findings in this study are openly available from the Dataverse repository at <https://doi.org/10.34894/XFZZPJ>.

Code availability

The Julia code to calculate the force response of the HWLC is openly available from the GitHub repository accessible at <https://doi.org/10.5281/zenodo.5970943>.

References

1. Paulson, J. R., Hudson, D. F., Cisneros-Soberanis, F. & Earnshaw, W. C. Mitotic chromosomes. *Semin. Cell Dev. Biol.* **117**, 7–29 (2021).
2. Hudson, D. F., Vagnarelli, P., Gassmann, R. & Earnshaw, W. C. Condensin is required for nonhistone protein assembly and structural

integrity of vertebrate mitotic chromosomes. *Dev. Cell* **5**, 323–336 (2003).

3. Samejima, K. et al. Mitotic chromosomes are compacted laterally by KIF4 and condensin and axially by topoisomerase II α . *J. Cell Biol.* **199**, 755–770 (2012).
4. Broedersz, C. P. & MacKintosh, F. C. Modeling semiflexible polymer networks. *Rev. Mod. Phys.* **86**, 995–1036 (2014).
5. Nielsen, C. F., Zhang, T., Barisic, M., Kalitsis, P. & Hudson, D. F. Topoisomerase II α is essential for maintenance of mitotic chromosome structure. *Proc. Natl Acad. Sci. USA* **117**, 12131–12142 (2020).
6. Nicklas, R. B. Measurements of the force produced by the mitotic spindle in anaphase. *J. Cell Biol.* **97**, 542–548 (1983).
7. Ye, A. A., Cane, S. & Maresca, T. J. Chromosome biorientation produces hundreds of piconewtons at a metazoan kinetochore. *Nat. Commun.* **7**, 13221 (2016).
8. Paulson, J. R. & Laemmli, U. K. The structure of histone-depleted metaphase chromosomes. *Cell* **12**, 817–828 (1977).
9. Farr, C. J., Antoniou-Kouroulioti, M., Mimmack, M. L., Volkov, A. & Porter, A. C. G. The α isoform of topoisomerase II is required for hypercompaction of mitotic chromosomes in human cells. *Nucleic Acids Res.* **42**, 4414–4426 (2014).
10. Hirano, T. & Mitchison, T. J. Topoisomerase II does not play a scaffolding role in the organization of mitotic chromosomes assembled in *Xenopus* egg extracts. *J. Cell Biol.* **120**, 601–612 (1993).
11. Earnshaw, W. C., Halligan, B., Cooke, C. A., Heck, M. M. S. & Liu, L. F. Topoisomerase II is a structural component of mitotic chromosome scaffolds. *J. Cell Biol.* **100**, 1706–1715 (1985).
12. Ono, T. et al. Differential contributions of condensin I and condensin II to mitotic chromosome architecture in vertebrate cells. *Cell* **115**, 109–

121 (2003).

13. Shintomi, K. & Hirano, T. Guiding functions of the C-terminal domain of topoisomerase II α advance mitotic chromosome assembly. *Nat. Commun.* **12**, 2917 (2021).
14. Neuman, K. C. & Nagy, A. Single-molecule force spectroscopy: optical tweezers, magnetic tweezers and atomic force microscopy. *Nat. Methods* **5**, 491–505 (2008).
15. Man, T., Witt, H., Peterman, E. J. G. & Wuite, G. J. L. The mechanics of mitotic chromosomes. *Q. Rev. Biophys.* **54**, e10 (2021).
16. Houchmandzadeh, B., Marko, J. F., Chatenay, D. & Libchaber, A. Elasticity and structure of eukaryote chromosomes studied by micromanipulation and micropipette aspiration. *J. Cell Biol.* **139**, 1–12 (1997).
17. Poirier, M., Eroglu, S., Chatenay, D. & Marko, J. F. Reversible and irreversible unfolding of mitotic newt chromosomes by applied force. *Mol. Biol. Cell* **11**, 269–276 (2000).
18. Sun, M., Kawamura, R. & Marko, J. F. Micromechanics of human mitotic chromosomes. *Phys. Biol.* **8**, 015003 (2011).
19. Sun, M., Biggs, R., Hornick, J. & Marko, J. F. Condensin controls mitotic chromosome stiffness and stability without forming a structurally contiguous scaffold. *Chromosome Res.* **26**, 277–295 (2018).
20. Gibcus, J. H. et al. A pathway for mitotic chromosome formation. *Science* **359**, eaao6135 (2018).
21. Walther, N. et al. A quantitative map of human condensins provides new insights into mitotic chromosome architecture. *J. Cell Biol.* **217**, 2309–2328 (2018).
22. Garcia-Exposito, L. et al. Proteomic profiling reveals a specific role for translesion DNA polymerase η in the alternative lengthening of telomeres. *Cell Rep.* **17**, 1858–1871 (2016).

23. Spector, D. L., Goldman, R. D. & Leinwand, L. A. *Cells: A Laboratory Manual* (Cold Spring Harbor Laboratory Press, 1998).
24. Heller, I., Hoekstra, T. P., King, G. A., Peterman, E. J. G. & Wuite, G. J. L. Optical tweezers analysis of DNA–protein complexes. *Chem. Rev.* **114**, 3087–3119 (2014).
25. Heller, I. et al. STED nanoscopy combined with optical tweezers reveals protein dynamics on densely covered DNA. *Nat. Methods* **10**, 910–916 (2013).
26. Schoen, I., Ries, J., Klotzsch, E., Ewers, H. & Vogel, V. Binding-activated localization microscopy of DNA structures. *Nano Lett.* **11**, 4008–4011 (2011).
27. Rubinstein, M. & Colby, R. H. *Polymer Physics* (Oxford Univ. Press, 2003).
28. Gardel, M. L. et al. Elastic behavior of cross-linked and bundled actin networks. *Science* **304**, 1301–1305 (2004).
29. Marko, J. F. & Siggia, E. D. Stretching DNA. *Macromolecules* **28**, 8759–8770 (1995).
30. Semmrich, C. et al. Glass transition and rheological redundancy in F-actin solutions. *Proc. Natl Acad. Sci. USA* **104**, 20199–20203 (2007).
31. Lieleg, O. & Bausch, A. R. Cross-linker unbinding and self-similarity in bundled cytoskeletal networks. *Phys. Rev. Lett.* **99**, 1–4 (2007).
32. Broedersz, C. P. et al. Measurement of nonlinear rheology of cross-linked biopolymer gels. *Soft Matter* **6**, 4120–4127 (2010).
33. Hliscs, R., Mühlig, P. & Claussen, U. The nature of G-bands analyzed by chromosome stretching. *Cytogenet. Cell Genet.* **79**, 162–166 (1997).
34. Kireeva, N., Lakonishok, M., Kireev, I., Hirano, T. & Belmont, A. S. Visualization of early chromosome condensation: a hierarchical

- folding, axial glue model of chromosome structure. *J. Cell Biol.* **166**, 775–785 (2004).
35. Adachi, Y., Luke, M. & Laemmli, U. K. Chromosome assembly in vitro: topoisomerase II is required for condensation. *Cell* **64**, 137–148 (1991).
 36. Johnson, M., Phua, H. H., Bennett, S. C., Spence, J. M. & Farr, C. J. Studying vertebrate topoisomerase 2 function using a conditional knockdown system in DT40 cells. *Nucleic Acids Res.* **37**, e98 (2009).
 37. Piskadlo, E. & Oliveira, R. A. A topology-centric view on mitotic chromosome architecture. *Int. J. Mol. Sci.* **18**, 2751 (2017).
 38. Poirier, M. G., Monhait, T. & Marko, J. F. Reversible hypercondensation and decondensation of mitotic chromosomes studied using combined chemical-micromechanical techniques. *J. Cell. Biochem.* **85**, 422–434 (2002).
 39. Strom, A. R. et al. HP1 α is a chromatin crosslinker that controls nuclear and mitotic chromosome mechanics. *eLife* **10**, e63972 (2021).
 40. Tamayo, J. Structure of human chromosomes studied by atomic force microscopy: part II. Relationship between structure and cytogenetic bands. *J. Struct. Biol.* **141**, 189–197 (2003).
 41. Hoshi, O. & Ushiki, T. Three-dimensional structure of G-banded human metaphase chromosomes observed by atomic force microscopy. *Arch. Histol. Cytol.* **64**, 475–482 (2001).
 42. Saldivar, J. C. et al. An intrinsic S/G2 checkpoint enforced by ATR. *Science* **361**, 806–810 (2018).
 43. Hochegger, H. et al. An essential role for Cdk1 in S phase control is revealed via chemical genetics in vertebrate cells. *J. Cell Biol.* **178**, 257–268 (2007).
 44. Dull, T. et al. A Third-generation lentivirus vector with a conditional packaging system. *J. Virol.* **72**, 8463–8471 (1998).

45. Rother, J., Nöding, H., Mey, I. & Janshoff, A. Atomic force microscopy-based microrheology reveals significant differences in the viscoelastic response between malign and benign cell lines. *Open Biol.* **4**, 140046 (2014).

Acknowledgements

We thank I. Samejima and W. Earnshaw for providing training in chromosome isolation; R. O’Sullivan for the U2OS TRF1-BirA cell line; S. Acar for help with cell culture; V. Bjerregaard for help with the lentiviral transduction; and A. Biebricher for help with optical tweezers. This work was supported by European Union Horizon 2020 grants (Chromavision 665233 to G.J.L.W., I.D.H., E.J.G.P. and Y.L.; and Antihelix 859853 to Y.L. and I.D.H.), the European Research Council under the European Union’s Horizon 2020 research and innovation program (MONOCHROME, grant agreement no. 883240 to G.J.L.W.), the Novo Nordisk Foundation (NNF18OC0034948 to I.D.H. and G.J.L.W.), the Deutsche Forschungsgemeinschaft (WI 5434/1-1 to H.W.), the Dutch Research Council (NWO Vidi 640-47-555 to I.H.), the Nordea Foundation (to I.D.H.) and the Danish National Research Foundation (DNRF115 to Y.L. and I.D.H.).

Author information

Author notes

1. These authors contributed equally: Anna E. C. Meijering, Kata Sarlós, Christian F. Nielsen
2. These authors jointly supervised this work: Erwin J. G. Peterman, Ian D. Hickson, Gijs J. L. Wuite

Affiliations

1. Department of Physics and Astronomy and LaserLaB Amsterdam, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands

Anna E. C. Meijering, Hannes Witt, Janni Harju, Emma Kerklingh, Guus H. Haasnoot, Iddo Heller, Chase P. Broedersz, Erwin J. G. Peterman & Gijs J. L. Wuite

2. Center for Chromosome Stability and Center for Healthy Aging,
Department of Cellular and Molecular Medicine, University of
Copenhagen, Copenhagen, Denmark

Kata Sarlós, Christian F. Nielsen, Anna H. Bizard, Ying Liu & Ian D. Hickson

Contributions

G.J.L.W., I.D.H., E.J.G.P., Y.L. and I.H. conceived the study. A.E.C.M., E.K., G.H.H. and H.W. performed the optical tweezers experiments. K.S., C.F.N. and A.H.B. developed cell lines and performed biochemical analysis of chromosomes. A.E.C.M., H.W., C.P.B. and G.J.L.W. performed data analysis and interpretation. J.H. and C.P.B developed the HWLC model. A.E.C.M., H.W., C.F.N., J.H., C.P.B., E.J.G.P., I.D.H. and G.J.L.W. wrote the manuscript. G.J.L.W., I.D.H., E.J.G.P and Y.L. led the research and the interpretation of the results. All authors discussed the results and commented on the manuscript.

Corresponding authors

Correspondence to [Chase P. Broedersz](#), [Ian D. Hickson](#) or [Gijs J. L. Wuite](#).

Ethics declarations

Competing interests

G.J.L.W., E.J.G.P. and I.H. own shares of LUMICKS. G.J.L.W. and E.J.G.P. serve on the technical advisory board of LUMICKS.

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 Biotin labelling, attachment and imaging of mitotic chromosomes.](#)

a, Representative images of chromosome spreads from U2OS cells stained with Alexa Fluor 568 Streptavidin, telomere repeat-binding factor 2 (TRF2) and DAPI to show an overlap between the Streptavidin foci and TRF2 staining at the telomeres. **b**, Representative bright-field images of chromosome attachment to two microspheres using flow-stretching. **c**, Approximately 20% of chromosomes became attached to the first microsphere in a ‘hugged’ conformation (along the length of a single sister chromatid), probably as a result of the binding of the telomeres from opposite ends of the chromatid to one microsphere. In these cases, the chromosome and bead would be rejected either immediately or following the inevitable failure to attach a second bead during attempted chromosome flow-stretching. **d**, Quantification of attachment efficiency of chromosomes to microspheres from wild-type U2OS cells and BirA-TRF1 expressing U2OS cells (mean values \pm s.e.m.) **e**, Representative images showing fluorescence of the intercalator YOPRO and a corresponding bright-field image. Small circular dark spots in panels **c** and **e** are a result of dust on optical surfaces outside of the flow cell. **f**, **g**, Representative images of immuno-staining of H3 histones (**f**) and NCAPH (**g**). **h**, Immuno-staining images for NCAPH and CREST to show localization of condensin I along the chromatids and one focus that reveals the position of a centromere. The

second centromere is not located in the focal plane and therefore is not visible.

Extended Data Fig. 2 Chromosome length, repeated stretching and stiffness calculation.

a, Histogram of stiffening length of U2OS and HCT116 chromosomes. **b**, Three consecutive elongation and retraction curves of a chromosome stretched to around 80 pN. **c**, Differential stiffness was determined by taking the numerical derivative of the force-extension data; that is, approximating the slope of a tangent line to the data. Stiffness-distance plots (inset) showing that the stiffness increases with distance.

Extended Data Fig. 3 Scheme of the HWLC.

Schematic depicting how the Hierarchical Worm-Like Chain (HWLC) captures the mechanical response of a chromosome as a function of force. The HWLC captures two distinct contributions of chromosome heterogeneity: 1) spatial heterogeneity (top row), where different chromosomal regions (depicted in red, orange and blue) exhibit distinct mechanical properties, and 2) scale-dependent heterogeneity (bottom row), where chromosomal structure has distinct mechanical properties on different length-scales (schematically depicted as red circles showing WLC elements of different scales, increasing stiffness at smaller scales is indicated by thicker lines) relating to different structural elements: examples shown here speculatively are chromatin loops and nucleosomes. These levels of chromosome heterogeneity are represented in the HWLC as a serial assembly of WLC elements with distinct linear stiffnesses and critical stiffening forces. The nonlinear mechanics of the HWLC emerges from the sequential stiffening of these underlying elements, reflecting how different chromosomal regions and length-scales dominate the mechanical response of the chromosome as the force is varied.

Extended Data Fig. 4 Individual and total response of WLC assemblies.

a, The individual and total responses for an assembly of three flexible WLCs. For this example, critical forces are given by a power-law; $\langle f \rangle_c \propto f_c^{-0.9}$, with cut-offs 0.0004 and 40 pN. All chains are of equal length $L = 1,000$ nm, and the constitutive law is $k_0 = L^{-1} f_c$. **b**, The individual and total response curves for an assembly of three flexible WLCs, with power-law distributed critical forces $\langle f \rangle_c \propto f_c^{-1}$, with cut-offs 0.0004 and 40 pN. The length of each component is chosen as $l(f) = 3,000 / (\sum_i f_{c,i}^{-2})^{1/2}$, so that the total length of the assembly is 3,000 nm, and the constitutive law is $k_0 \langle f \rangle_c = l \langle f \rangle_c^{-1} \propto f_c^{-3}$. **c**, The individual and total response curves for an assembly of 100 flexible WLCs with power-law distributed critical forces; $\langle f \rangle_c \propto f_c^{-\beta}$, $\beta = 0.85$, and equal component lengths; $k_0 \langle f \rangle_c = l \langle f \rangle_c^{-\alpha}$, $\alpha = 1$, $l = 30$ nm. Slope prediction is given by $\beta + \alpha - 1 = 0.85$. Cut-offs for $\langle f \rangle_c$ are 0.0004 and 40 pN. **d**, The individual and total response curves for an assembly of 100 semi-flexible WLCs with an exponential distribution of critical forces: $\langle f \rangle_c = e^{-f_c/50}/50$. The linear spring coefficients are $k_0 = f_c^2 \times 90 / (\pi k_B T)$.

Extended Data Fig. 5 Effect of model parameters on the HWLC.

The panels in each vertical column show the effect of varying one model parameter (the number of sub-chains, N , the chromosome length, L , or the lower cut-off for the critical force, $f_{c,\min}$) on the distribution of stiffness curves for the power-law (**a–c**) and exponential distribution (**d–f**). Other parameters are kept at the same values as in the middle row. **a, d**, The number of sub-chains affects the initial stiffness of the assembly and sets the scale of the variance of stiffnesses. **b, e**, The stiffness of the assembly varies inversely with its total length. **c, f**, The lower cut-off for the distribution of critical forces sets the initial stiffness of the ensemble and the force at which it starts to stiffen. Line and shaded regions depict mean and standard deviation of experimental data of U2OS chromosomes (grey), HWLC model

with $\langle P(f_c) \rangle_{proto} \{f_c\}^{\{-\beta\}}$ (orange) and HWLC model with $\langle P(f_c) \rangle_{proto} \{rm\exp\}(-f_c/\{f_c\}^{*\})$ (yellow).

Extended Data Fig. 6 Characterization of cell synchronization and TOP2A depletion.

a, Immunoblot of wild-type HCT116 cells (WT) and two clones of CDK1as-modified cells, in which the genes encoding the 35 kDa endogenous hsCDK1 has been knocked out (top panel) and the mutant CDK1as gene has been knocked-in and is expressed ectopically (*xl*CDK1as; 40 kDa; middle panel). *xl*CDK1as is not recognized by the hsCDK1 antibody and was detected using an antibody to the c-myc epitope tag. #1 is a TOP2A-mAID clone and #2 the TOP2A-mAID H2B-EGFP clone used in this study. Histone H3 was used as a loading control (bottom panel). For gel source data, see Supplementary Fig. 1. **b**, Optimization of CDK1as cell synchronization. Quantification of G1 and G2/M content from Propidium Iodide-flow cytometry profiles of untreated HCT116 TOP2A-mAID CDK1as cells, HCT116 TOP2A-mAID CDK1as cells grown with 0.25 μ M 1NM-PP1 for 16 h to arrest them in G2 phase, and HCT116 TOP2A-mAID CDK1as cells grown with 0.25 μ M 1NM-PP1 for 16 h and then released from this block and grown for 4 h. **c**, Diagram of the experimental procedure for TOP2A depletion. **d**, Quantification of mitotic cells at the indicated time points following release from CDK1as arrest, as depicted in (c). **e**, Representative immunofluorescent images of cells released from CDK1as arrest at the indicated timepoints shown in (d). pH3S10 antibody (red) was used as a mitotic marker and DAPI (blue) was used to stain DNA. **f**, Quantification of chromosome morphology into four categories (based on degree of condensation) from chromosome spreads in control versus TOP2A degraded cells with representative images of the categories on the right (error bars: s.e.m.). Values are the mean of three independent experiments with a total count as mentioned in the figure. **g**, Representative images of immunostaining for TOP2A on chromosome spreads of control and the auxin-exposed HCT116 TOP2A-mAID CDK1as cells to confirm degradation of TOP2A in the auxin-exposed chromosomes.

Extended Data Fig. 7 Additional characterization of the mechanics of TOP2A-depleted chromosomes.

a–c, Histograms of stiffening length (**a**), compliance at 200 pN (**b**) and strain stiffening exponent (**c**) for control and auxin-exposed (TOP2A-depleted) chromosomes. **d**, Kymograph of fluorescence signal during stretching of hypo-condensed, H2B-eGFP-labelled, TOP2A-depleted chromosomes. Above the kymograph, the initial position of the chromosome between microspheres is schematically illustrated. The boundaries of regions with different histone intensity signal are depicted by lines. **e**, Stiffening exponent of force-extension curves for HCT116 control chromosomes, auxin-exposed chromosomes and auxin-exposed chromosomes with a stiffening length above 5 μm before (B) and after (A) salt-induced decompaction. Two-sided Wilcoxon rank-sum test, * $p < 0.05$; ** $p < 0.01$. Centre: median, box: 25th to 75th percentile, whiskers: minimum and maximum datapoints (not considered as outlier). **f**, Force-extension curve of an HCT116 chromosome before salt-induced decompaction (maintained at 80 mM KCl; black line) and three consecutive force-extension curves during salt induced decompaction at 480 mM KCl.

Supplementary information

Supplementary Information

This file contains Supplementary Notes 1 and 2, and Supplementary Figure 1. The Supplementary Notes contain a detailed description of the HWLC model and a discussion of selection bias on condensed chromosomes. Supplementary Figure 1 contains raw images of gels.

Reporting Summary

Peer Review File

Supplementary Video 1

Representative video from bright-field images of chromosome attachment to two microspheres using flow stretching.

Supplementary Video 2

Representative video from bright-field images of stretching a U2OS chromosome.

Supplementary Video 3

Representative video from bright-field images of three consecutive extension and retraction cycles of a U2OS chromosome.

Supplementary Video 4

Representative video from bright-field images of consecutive extension and retraction cycles of a HCT116 control chromosome. The first two extensions are at 80 mM KCl, after which the KCl concentration is alternated between 480 mM and 80 mM.

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

Meijering, A.E.C., Sarlós, K., Nielsen, C.F. *et al.* Nonlinear mechanics of human mitotic chromosomes. *Nature* **605**, 545–550 (2022).
<https://doi.org/10.1038/s41586-022-04666-5>

- Received: 14 July 2021
- Accepted: 21 March 2022
- Published: 04 May 2022
- Issue Date: 19 May 2022
- DOI: <https://doi.org/10.1038/s41586-022-04666-5>

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/s41586-022-04666-5>

[Download PDF](#)

- Article
- Open Access
- [Published: 24 March 2022](#)

Design of protein-binding proteins from the target structure alone

- [Longxing Cao](#)^{1,2} na1,
- [Brian Coventry](#)^{1,2,3} na1,
- [Inna Goreshnik](#)^{1,2},
- [Buwei Huang](#)^{1,2,4},
- [William Sheffler](#)^{1,2},
- [Joon Sung Park](#) ORCID: orcid.org/0000-0002-1575-1849⁵,
- [Kevin M. Jude](#) ORCID: orcid.org/0000-0002-3675-5136^{6,7,8},
- [Iva Marković](#)^{9,10},
- [Rameshwar U. Kadam](#)¹¹,
- [Koen H. G. Verschueren](#)^{9,10},
- [Kenneth Verstraete](#)^{9,10},
- [Scott Thomas Russell Walsh](#)^{12,13},
- [Nathaniel Bennett](#)^{1,2,3},
- [Ashish Phal](#) ORCID: orcid.org/0000-0001-8405-4150^{1,4,14},
- [Aerin Yang](#) ORCID: orcid.org/0000-0002-8704-0288^{6,7,8},
- [Lisa Kozodoy](#)^{1,2},
- [Michelle DeWitt](#)^{1,2},
- [Lora Picton](#) ORCID: orcid.org/0000-0002-0482-5187^{6,7,8},
- [Lauren Miller](#)^{1,2},
- [Eva-Maria Strauch](#) ORCID: orcid.org/0000-0001-7382-747X¹⁵,
- [Nicholas D. DeBouver](#) ORCID: orcid.org/0000-0003-2545-6007^{16,17},
- [Allison Pires](#)^{17,18},
- [Asim K. Bera](#) ORCID: orcid.org/0000-0001-9473-2912^{1,2},

- [Samer Halabiya¹⁹](#),
- [Bradley Hammerson¹⁷](#),
- [Wei Yang^{1,2}](#),
- [Steffen Bernard¹¹](#),
- [Lance Stewart](#) ORCID: [orcid.org/0000-0003-4264-5125^{1,2}](https://orcid.org/0000-0003-4264-5125),
- [Ian A. Wilson](#) ORCID: [orcid.org/0000-0002-6469-2419^{11,20}](https://orcid.org/0000-0002-6469-2419),
- [Hannele Ruohola-Baker](#) ORCID: [orcid.org/0000-0002-5588-4531^{1,14}](https://orcid.org/0000-0002-5588-4531),
- [Joseph Schlessinger](#) ORCID: [orcid.org/0000-0002-5085-5969⁵](https://orcid.org/0000-0002-5085-5969),
- [Sangwon Lee⁵](#),
- [Savvas N. Savvides](#) ORCID: [orcid.org/0000-0003-3420-5947^{9,10}](https://orcid.org/0000-0003-3420-5947),
- [K. Christopher Garcia](#) ORCID: [orcid.org/0000-0001-9273-0278^{6,7,8}](https://orcid.org/0000-0001-9273-0278) &
- [David Baker](#) ORCID: [orcid.org/0000-0001-7896-6217^{1,2,21}](https://orcid.org/0000-0001-7896-6217)

Nature volume **605**, pages 551–560 (2022)

- 63k Accesses
- 3 Citations
- 315 Altmetric
- [Metrics details](#)

Subjects

- [High-throughput screening](#)
- [Protein design](#)

Abstract

The design of proteins that bind to a specific site on the surface of a target protein using no information other than the three-dimensional structure of the target remains a challenge^{1,2,3,4,5}. Here we describe a general solution to this problem that starts with a broad exploration of the vast space of possible binding modes to a selected region of a protein surface, and then intensifies the search in the vicinity of the most promising binding modes. We

demonstrate the broad applicability of this approach through the de novo design of binding proteins to 12 diverse protein targets with different shapes and surface properties. Biophysical characterization shows that the binders, which are all smaller than 65 amino acids, are hyperstable and, following experimental optimization, bind their targets with nanomolar to picomolar affinities. We succeeded in solving crystal structures of five of the binder–target complexes, and all five closely match the corresponding computational design models. Experimental data on nearly half a million computational designs and hundreds of thousands of point mutants provide detailed feedback on the strengths and limitations of the method and of our current understanding of protein–protein interactions, and should guide improvements of both. Our approach enables the targeted design of binders to sites of interest on a wide variety of proteins for therapeutic and diagnostic applications.

Main

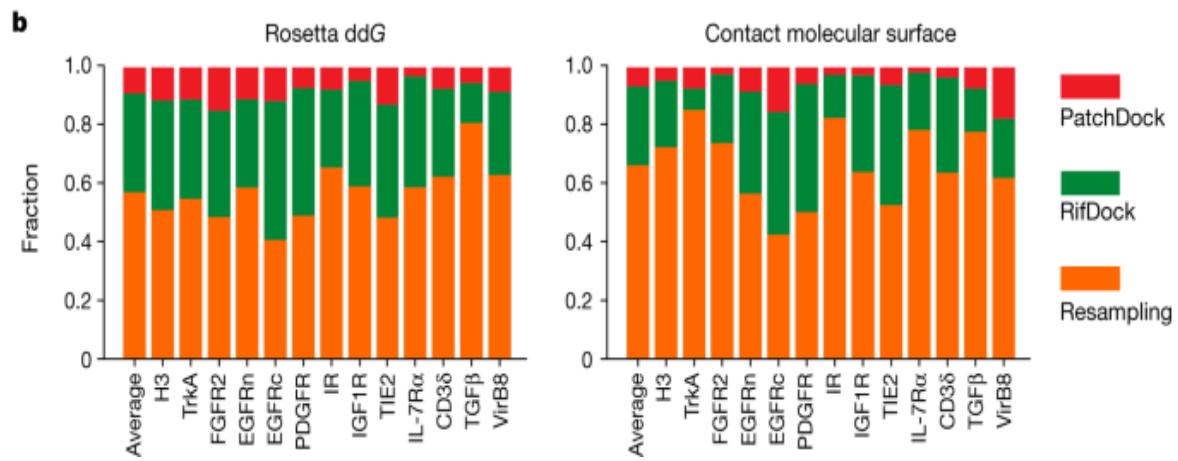
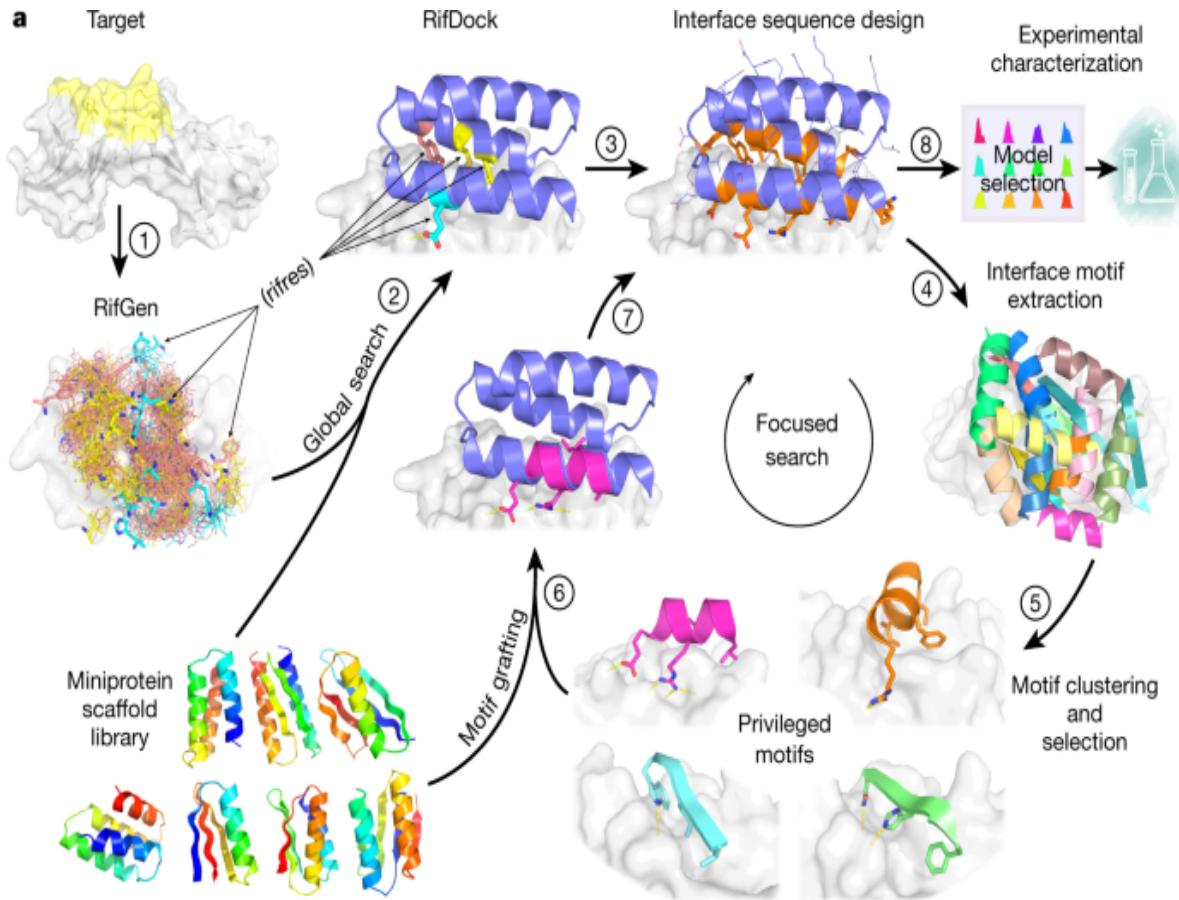
Protein interactions have crucial roles in biology, and general approaches to design proteins that disrupt or modulate these interactions would have great utility. Empirical selection approaches that start from large antibody, designed ankyrin repeat protein or other protein scaffold libraries can generate binders to protein targets. However, it is difficult at the outset to target a specific region on a target protein surface and to sample the entire space of possible binding modes. Computational methods can target specific target surface locations and provide a more principled and a potentially faster approach to generate binders than random library selection methods, as well as insight into the fundamental properties of protein interfaces (which must be understood for design to be successful). Most current computational methods used to design proteins that bind to a target surface utilize information derived from structures of the native complex on specific side-chain interactions or protein backbone placements optimal for binding^{1,2,3}. Computational docking of antibody scaffolds with varied loop geometries has yielded binders, but the designed binding modes have rarely been validated with high-resolution structures⁴. Binders have been generated starting from several computationally identified hotspot residues, which were then used to guide the positioning of naturally occurring protein scaffolds⁵. However, for many target proteins, there are no obvious pockets

or clefts on the protein surface into which a small number of privileged side chains can be placed, and guidance by a small number of hotspot residues limits the approach to a small fraction of possible interaction modes.

Design method

We sought to develop a general approach to the design of high-affinity binders to arbitrary protein targets that addresses two major challenges. First, there are generally no clear side-chain interactions or secondary structure packing arrangements that can mediate strong interactions with the target; instead there are vast numbers of individually very weak possible interactions. Second, the number of ways of choosing which of these numerous weak interactions to incorporate into a single binding protein is combinatorially large, and any given protein backbone is unlikely to be able to simultaneously present side chains that can encompass any preselected subset of these interactions. To illustrate our approach, consider the simple analogy of a difficult climbing wall with only a few suitable footholds or handholds distant from each other. Previous hotspot-based approaches correspond to focusing on routes that involve these footholds and handholds, but this greatly limits possibilities and there may be no way to connect them into a successful route. An alternative is to first identify all the possible handholds and footholds, no matter how poor; second, have thousands of climbers select subsets of these and try to climb the wall; third, identify those routes that showed the most promise, and fourth, have a second group of climbers explore them in detail. Following this analogy, we devised the following multistep approach to overcome the above two challenges: step (1), enumerate a large and comprehensive set of disembodied side-chain interactions with the target surface; step (2), identify from large *in silico* libraries of protein backbones those that can host many of these side chains without clashing with the target; step (3), identify recurrent backbone motifs in these structures; and step (4), generate and place against the target a second round of scaffolds that contain these interacting motifs (Fig. 1a and Extended Data Fig. 1). Steps (1) and (2) widely search the space, whereas steps (3) and (4) intensify the search in the regions that show the most promise. We describe each step in more detail below.

Fig. 1: Overview of the de novo protein binder design pipeline.



a, Schematic of our two-stage binder design approach. In the global search stage, billions of disembodied amino acids are docked onto the selected region of the target protein surface using RifGen, the favourable interacting amino acids are stored as rifres (step 1), and miniprotein scaffolds are then docked on the target guided by these favourable side-chain interactions (step 2). The interface sequences are then designed to maximize interactions with

the target (step 3). In the focused search stage, interface structural motifs are extracted and clustered (steps 4 and 5). These privileged motifs are then used to guide another round of docking and design (steps 6 and 7). Designs are then selected for experimental characterization based on computational metrics (step 8). See Extended Data Fig. 1 for a more detailed flow chart of the de novo binder design pipeline. **b**, Comparison of the sampling efficiency of PatchDock, RifDock and resampling protocols. Bar graph shows the distribution over the three approaches of the top 1% of binders based on Rosetta ddG and contact molecular surface values after pooling equal-CPU-time dock-and-design trajectories for each of the 13 target sites and averaging per-target distributions ([Methods](#)).

We began by docking disembodied amino acids against the target protein and storing the backbone coordinates and target binding energies of the typically billions of amino acids that make favourable hydrogen bonding or nonpolar interactions in a six-dimensional spatial hash table for rapid look-up (Fig. 1a and [Methods](#)). This rotamer interaction field (RIF) enables rapid approximation of the target interaction energy achievable by a protein scaffold docked against a target based on its backbone coordinates alone (with no need for time-consuming side-chain sampling). For each dock, the target interaction energies of each of the matching amino acids in the hash table are summed. A related approach was used for the design of small-molecule binders⁶; as protein targets are so much bigger and because nonpolar interactions are the primary driving force for protein–protein interactions, we focused the RIF generation process on nonpolar sites in specific surface regions of interest. For example, for the design of inhibitors, we focused on interaction sites with biological partners. The RIF approach improves on previous discrete interaction-sampling approaches⁵ by reducing the algorithmic complexity from $O(N)$ or $O(N^2)$ to $O(1)$ with respect to the number of side-chain–target interactions considered, thereby allowing for billions, rather than thousands, of potential interfaces to be considered.

For docking against the RIF, it is desirable to have a large set of protein scaffold options, as the chance that any one scaffold can house many interactions is small. The structure models of these scaffolds must be quite accurate so that the positioning is correct. Using fragment assembly⁷, piecewise fragment assembly⁸ and helical extension⁹, we designed a large

set of miniproteins that ranged in length from 50 to 65 amino acids and contained larger hydrophobic cores than previous miniprotein scaffold libraries¹. These properties make the protein more stable and more tolerant to the introduction of the designed binding surfaces. A total of 84,690 scaffolds spanning 5 different topologies with structural metrics predictive of folding were encoded in large oligonucleotide arrays, and 34,507 of these were found to be stable using a high-throughput proteolysis-based protein stability assay¹⁰.

We experimented with several approaches for docking these stable scaffolds against the target structure RIF, balancing overall shape complementarity with maximizing specific rotamer interactions. The most robust results were obtained using direct low-resolution shape matching¹¹ followed by grid-based refinement of the rigid body orientation in the RIF (RIFDock). This approach resulted in better Rosetta binding energy (ddG) values and packing (contact molecular surface, see below) after sequence design than shape matching alone with PatchDock (Fig. 1b, red and green), and more extensive nonpolar interactions with the target than hierarchical search without PatchDock shape matching (Extended Data Fig. 2a). ⁶

Because of the loss in resolution in the hashing used to build the RIF, and the necessarily approximate accounting for interactions between side chains ([Methods](#)), we found that evaluation of the RIF solutions was considerably enhanced by full combinatorial optimization using the Rosetta forcefield, which allow the target side chains to repack and the scaffold backbone to relax. However, full combinatorial sequence optimization is CPU intensive. To enable efficient screening of millions of alternative backbone placements, we developed a rapid interface pre-screening method using Rosetta to identify promising RIF docks. Restricting to hydrophobic amino acids and considering a smaller number of side-chain rotamers than in standard Rosetta design calculations, together with a more rapidly computable energy function sped up the design time by more than tenfold while retaining a strong correlation with results after full sequence design (next paragraph). This pre-screen (referred to as the ‘Predictor’ below) substantially improved the binding energies and shape complementarity of the final designs, as far more RIF solutions could be processed (Extended Data Fig. 2b).

We observed that application of the standard Rosetta design to the set of filtered docks in some cases resulted in models with buried unsatisfied polar groups and other suboptimal properties. To overcome these limitations, we developed a combinatorial sequence design protocol that maximizes shape and chemical complementarity with the target while avoiding buried polar atoms. Sequence compatibility with the scaffold monomer structure was increased using a structure-based sequence profile¹², cross-interface interactions were upweighted during the Monte-Carlo-based sequence design stage to maximize the contacts between the binder and the target (ProteinProteinInterfaceUpweighter; [Methods](#)) and rotamers that contained buried unsatisfiable polar atoms were eliminated before packing and buried unsatisfied polar atoms penalized by a pair-wise decomposable pseudo-energy term¹³. This protocol yielded amino acid sequences that were more strongly predicted to fold to the designed structure (Extended Data Fig. [2c](#)) and to bind the target (Extended Data Fig. [2d](#)) than standard Rosetta interface design.

In the course of developing the overall binder design pipeline, we noted after inspection that even designs with favourable Rosetta binding free energies, large changes in the solvent-accessible surface area (SASA) after binding and high shape complementarity (SC) often lacked dense packing and interactions that involve several secondary structural elements. We developed a quantitative measure of packing quality in closer accord with visual assessment—the contact molecular surface ([Methods](#))—which balances interface complementarity and size in a manner that explicitly penalizes poor packing. We used this metric to help to select suitable designs at both the rapid Predictor stage and after full sequence optimization ([Methods](#)).

The space sampled by the search across the structure and sequence space is enormous: tens of thousands of possible protein backbones \times nearly 1 billion possible disembodied side-chain interactions per target $\times 10^{16}$ interface sequences per scaffold placement. Sampling of spaces of this size is necessarily incomplete, and many of the designs at this stage contained buried unsatisfied polar atoms (only rotamers that cannot make hydrogen bonds in any context are excluded at the packing stage) and cavities. To generate improved designs, we intensified the search around the best of the

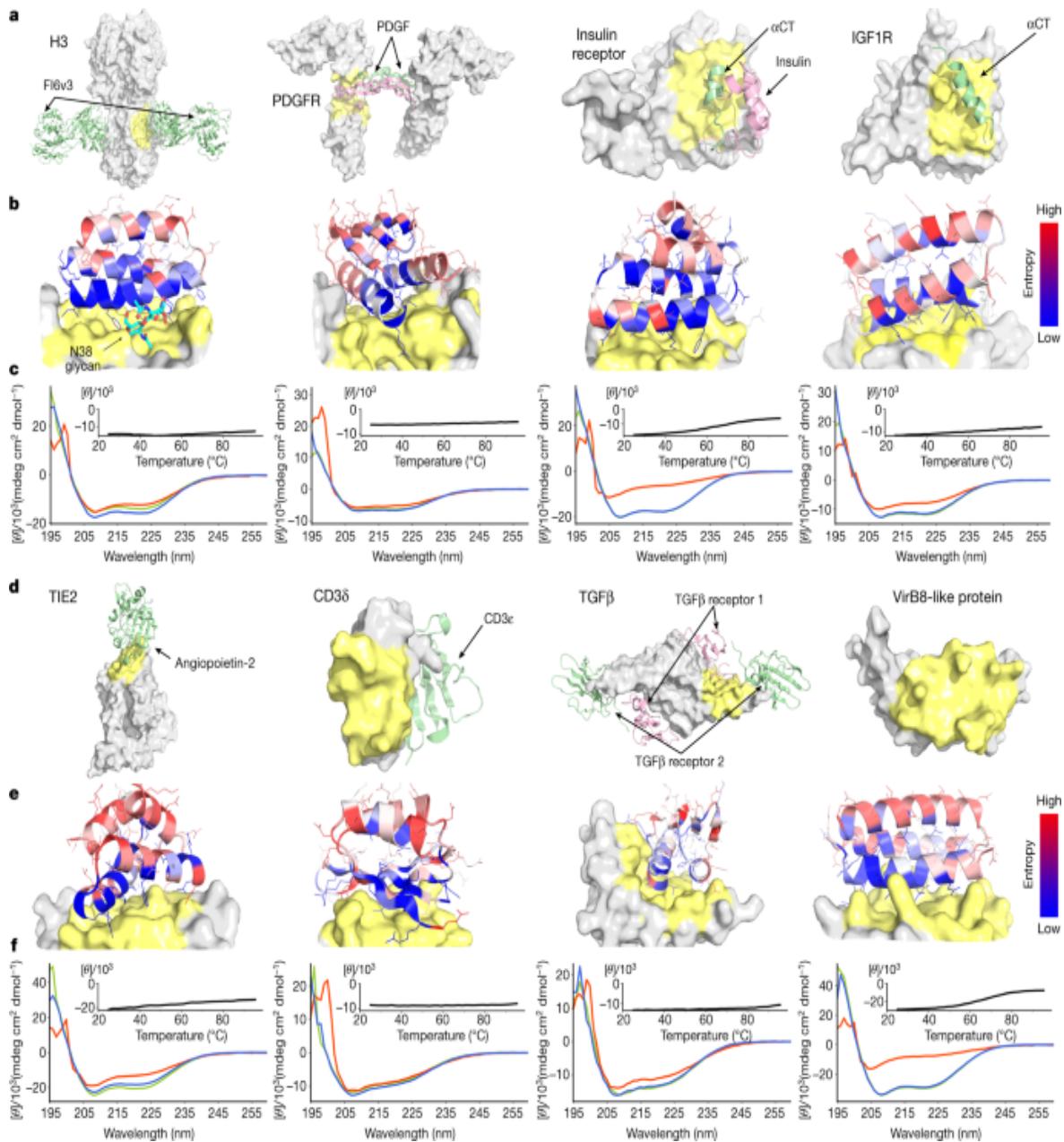
designed interfaces. We developed a resampling protocol that first extracts all the secondary structural motifs that make good contacts with the target protein from the first ‘broad search’ designs. Next, it clusters these motifs on the basis of their backbone coordinates and rigid body placements, and then selects the binding motif in each cluster with the best per-position weighted Rosetta binding energy. Using this approach, around 2,000 motifs were selected for each target. These motifs, which in many cases resemble canonical secondary structure packing patterns¹⁴, are privileged because they contain a much greater density of favourable side-chain interactions with the target than the rest of the designs. The motifs were used to guide another round of docking and design. First, scaffolds from the library were superimposed on the motifs and the favourable-interacting motif residues transferred to the scaffold. The remainder of the scaffold sequence was optimized to make further interactions with the target, allowing backbone flexibility through backbone torsion-angle minimization to increase shape complementarity with the target (Fig. 1a). Design Interface metrics following resampling were considerably improved over those from the broad searching stage (Fig. 1b). The designs with the most favourable protein folding and protein interface metrics from both the broad searching and resampling stages were selected for experimental validation.

Experimental testing

Previous approaches used to design protein binders have been tested on only one or two targets, which limits assessment of their generality. To thoroughly test our new binder design pipeline, we selected 13 native proteins of considerable current interest and spanning a wide range of shapes and biological functions. These proteins fall into two classes: (1) human cell surface or extracellular proteins involved in signalling, and (2) pathogen surface proteins. Binders for human cell surface or extracellular proteins could have utility as probes of biological mechanism and potentially as therapeutics, and hence we sought to design binders to tropomyosin receptor kinase A (TrkA; also known as NTRK1)¹⁵, fibroblast growth factor receptor 2 (FGFR2)¹⁶, epidermal growth factor receptor (EGFR)¹⁷, platelet-derived growth factor receptor (PDGFR)¹⁸, insulin receptor (IR)¹⁹, insulin-like growth factor 1 receptor (IGF1R)²⁰, angiopoietin-1 receptor (TIE2)²¹,

interleukin-7 receptor- α (IL-7R α)²², CD3 delta chain (CD3 δ)²³ and transforming growth factor- β (TGF β)²⁴. Binding proteins for pathogen surface proteins could also have therapeutic utility, and so we also designed binders to influenza A H3 haemagglutinin (H3)²⁵, VirB8-like protein from *Rickettsia typhi* (VirB8)²⁶ and the SARS-CoV-2 coronavirus spike protein (Figs. 2 and 3). For each of these surface proteins, we selected one or two regions for the binders to target to ensure maximal biological utility and for potential downstream therapeutic potential. These regions span a wide range of surface properties, with diverse shape and chemical characteristics (Figs. 2 and 3, and Extended Data Fig. 3). Some of the selected targeting regions overlap with the native interfaces, but no native interface information or native hotspots were used during the binder design process. For some targets (for example, CD3 δ and VirB8), no structures of the native complex were available and there were no proteins known to bind at the targeted region.

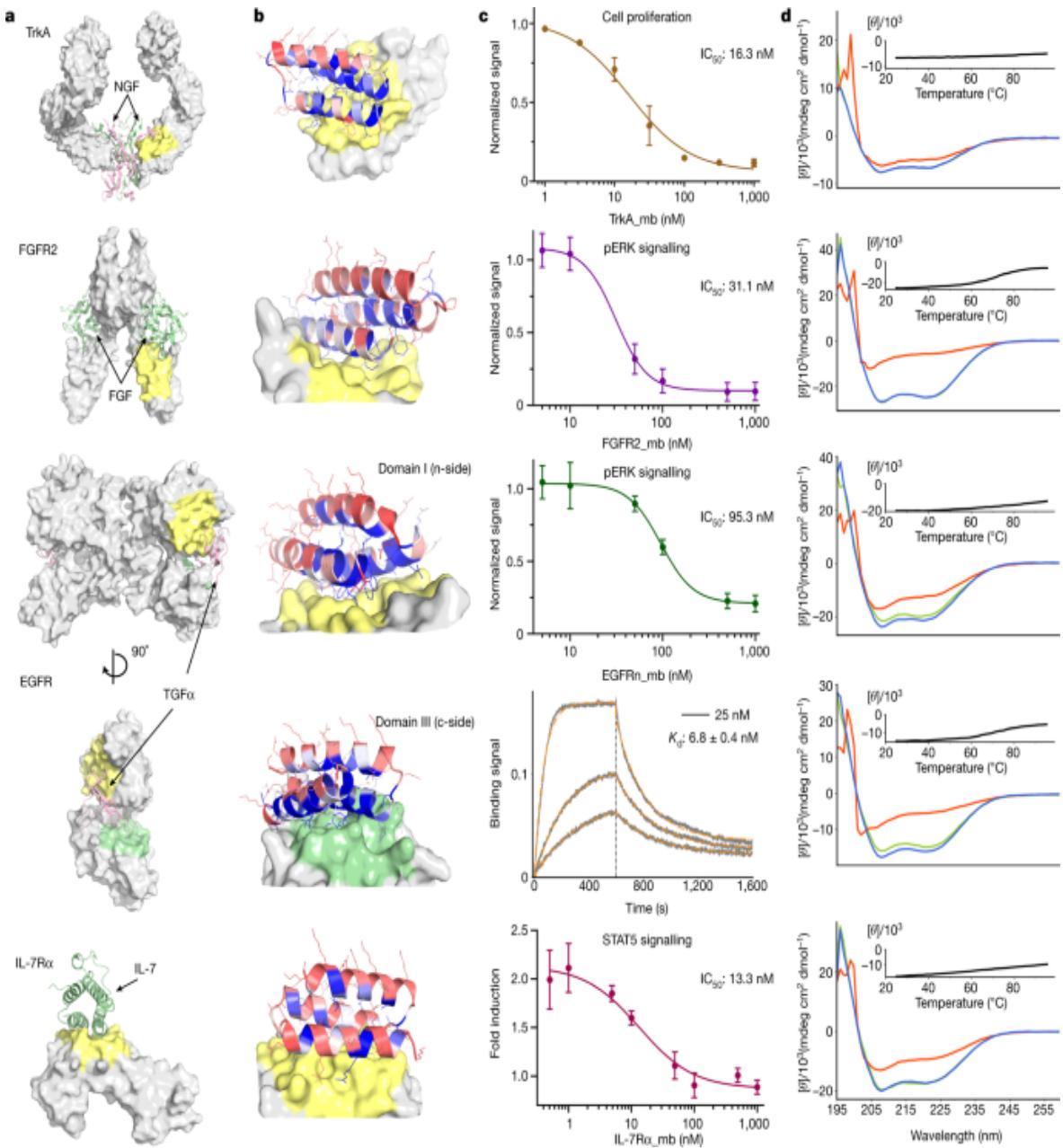
Fig. 2: De novo design and characterization of miniprotein binders.



a, d, Naturally occurring target protein structures shown in surface representation, with known interacting partners in cartoons where available. Regions targeted for binder design are coloured in pale yellow or green; the remainder of the target surface is in grey. See Extended Data Fig. 3 for side-by-side comparisons of the native binding partners and the computational design models. The PDB identifiers are [3ZTJ](#) (H3), [3MJG](#) (PDGFR), [4OGA](#) (IR), [5U8R](#) (IGF1R), [2GY7](#) (TIE2), [1XIW](#) (CD3 δ), [3KFD](#) (TGF β) and [4O3V](#) (VirB8). α CT, α -chain C-terminal helix. **b, e**, Computational models of designed complexes coloured by site saturation mutagenesis results.

Designed binding proteins (cartoons) are coloured by positional Shannon entropy, with blue indicating positions of low entropy (conserved) and red those of high entropy (not conserved); the target surface is in grey and yellow. The core residues and binding interface residues are more conserved than the non-interface surface positions, consistent with the computational models. Full SSM maps over all positions of all the de novo designs are provided in the [Supplementary Information](#). **c, f**, Circular dichroism spectra at different temperatures (green, 25 °C; red, 95 °C; blue, 95 °C followed by 25 °C), and circular dichroism signals at 222-nm wavelength as a function of temperature for the optimized designs (insets). See Extended Data Fig. 4 for the biolayer interferometry characterization results of the optimized designs.

Fig. 3: De novo design and inhibition of native signalling pathways by designed miniproteins.



See the panel descriptions in Fig. 2 legend for **a**, **b**, **d**. The PDB identifiers are [2IFG](#) (TrkA), [1DIS](#) (FGFR2), [1MOX](#) (EGFR) and [3DI3](#) (IL-7R α) for **a**. **c**, For TrkA, the dose-dependent reduction in cell proliferation after 48 h of TF-1 cells with increasing TrkA minibinder (TrkA_mb) concentration is shown. (8.0 ng ml⁻¹ human β -NGF was used for competition). Titration curves at different concentrations of NGF and the effects of the miniprotein binders on cell viability are presented in Extended Data Fig. 8. For FGFR2, the dose-dependent reduction pERK signalling elicited by 0.75 nM β -FGF in human umbilical vein endothelial cells (HUVECs) with increasing FGFR2

minibinder (FGFR2_mb) concentration is shown. For the EGFRn-side binder, the dose-dependent reduction in pERK signalling elicited by 1 nM EGF in HUVECs with increasing EGFRn-side minibinder (EGFRn_mb) concentration is shown. See Extended Data Fig. 9 and Methods for experimental details. For the EGFRc-side binder, biolayer interferometry results are shown. See Extended Data Fig. 4 for the biolayer interferometry characterization results of the other optimized designs. For IL-7R, the reduction in STAT5 activity induced by 50 pM of IL-7 in HEK293T cells in the presence of increasing IL-7Ra minibinder (IL-7Ra_mb) concentrations is shown. The mean values were calculated from triplicates for the cell signalling inhibition assays measured in parallel, and error bars represent standard deviations. IC₅₀ was calculated using a four-parameter-logistic equation in GraphPad Prism 9 software.

Using the above protocol, we designed 15,000–100,000 binders for each of the 13 target sites on the 12 native proteins ([Methods](#); we chose two sites for EGFR). Synthetic oligonucleotides (230 base pairs) encoding the 50–65 residue designs were cloned into a yeast surface-expression vector so that the designs were displayed on the surface of yeast. Those that bound their target were enriched by several rounds of fluorescence-activated cell sorting (FACS) using fluorescently labelled target proteins. The starting and enriched populations were deep sequenced, and the frequency of each design in the starting population and after each sort was determined. From multiple sorting rounds at different target protein concentrations, we determined, as a proxy for the binding dissociation constant (K_d) values, the midpoint concentration (SC₅₀) in the binding transitions for each design in the library (Extended Data Table 1 and [Methods](#)).

To assess whether the top enriched designs for each target fold and bind as in the corresponding computational design models, and to investigate the sequence dependence of folding and binding, we generated high-resolution footprints of the binding surface by sorting site saturation mutagenesis libraries (SSMs) in which every residue was substituted with each of the 20 amino acids one at a time. For the majority of the enriched designs, substitutions at the binding interface and in the protein core were less tolerated than substitutions at non-interface surface positions (Figs. 2 and 3, and Extended Data Fig. 5), and all of the cysteine residues were highly

conserved in designs that contained disulfides. The effects of each mutation on both binding energy and monomer stability were estimated using Rosetta design calculations, and a reasonable correlation was found between the predicted and experimentally determined effect of mutations (Extended Data Fig. 6a). In almost all cases, a small number of substitutions increased the apparent binding affinity, and we generated libraries combining 5–15 of these and sorted them for binding under increasingly stringent (lower target concentration) conditions. Many of these affinity-enhancing substitutions were mutations to tyrosine (Extended Data Fig. 6b), which is consistent with the relatively high frequency of tyrosine in natural protein interfaces²⁷. The set of affinity-increasing substitutions provide valuable information to improve the binder design approach, as these substitutions ideally would have been identified in the computational sequence design calculations (see ‘Discussion’ for more details).

We expressed the highest affinity combinatorially optimized binders for each target in *Escherichia coli* to enable more detailed structural and functional characterization. All of the designs were in the soluble fraction and could be readily purified by Ni²⁺-NTA chromatography. All had circular dichroism spectra consistent with the design model, and most (9 out of 13) were stable at 95 °C (Figs. 2 and 3, and Table 1). The binding affinities for the targets were assessed by biolayer interferometry and values ranged from 300 pM to 900 nM (Fig. 3, Table 1 and Extended Data Fig. 4). The sequence mapping data report on the residues in the design that are crucial for binding, but only weakly on the region of the target bound. We investigated the latter using a combination of binding competition experiments, biological assays and structural characterization of the complexes. For the nine targets for which these were available, this characterization suggested binding modes consistent with the design models, as described in the subsequent paragraphs.

Table 1 Physicochemical properties of the optimized de novo miniprotein binders

Cell receptors involved in signalling

The receptor tyrosine kinases TrkA, FGFR2, PDGFR, EGFR, IR, IGF1R and TIE2 are key regulators of cellular processes and are involved in the development and progression of many types of cancer²⁸. We designed binders that targeted the native ligand-binding sites for PDGFR, EGFR (on both domain I and domain III; the binders are referred to as EGFRn_mb and EGFRc_mb, respectively), IR, IGF1R and TIE2, whereas for TrkA and FGFR2, we targeted surface regions proximal to the native ligand-binding sites (Figs. 2 and 3; see [Methods](#) for criteria). We obtained binders to all eight target sites, and the binding affinities of the optimized designs ranged from about 1 nM or better for TrkA and FGFR2 to 860 nM for IGF1R (Table 1). Competition experiments with nerve growth factor (NGF), platelet-derived growth factor-BB (PDGF-BB), insulin, insulin growth factor 1 (IGF1) and angiopoietin 1 (ANG1) on yeast indicated that the binders for TrkA, PDGFR, IR, IGF1R and TIE2 bind to the targeted sites (Extended Data Fig. 7), consistent with the computational design models. The receptor tyrosine kinase binders are monomers, and as such are all expected to be antagonists. We tested the effect of the cognate binders on signalling through TrkA, FGFR2 and EGFR in cultured cells. Strong inhibition of signalling by the native agonists was observed in all three cases (Fig. 3c, and Extended Data Figs. 8 and 9).

Binding of IL-7 to the IL-7 α receptor subunit leads to recruitment of the γ_c receptor, which forms a tripartite cytokine–receptor complex crucial to signalling cascades that lead to the development and homeostasis of T and B cells²⁹. We obtained a picomolar affinity binder for IL-7R α targeting the IL-7 binding site and found that it blocks STAT5 signalling induced by IL-7 (Fig. 3c and Table 1). We also obtained binders to CD3 δ , one of the subunits of the T cell receptor, and the signalling molecule TGF β , which play pivotal parts in immune cell development and activation (Fig. 2 and Table 1).

Pathogen target proteins

Influenza haemagglutinin (HA) is the main target for influenza A virus vaccines and drugs, and can be genetically classified into two main subgroups: group 1 and group 2 (refs. [30,31](#)). The HA stem region is an attractive therapeutic epitope as it is highly conserved across all influenza A subtypes, and targeting this region can block the low-pH-induced

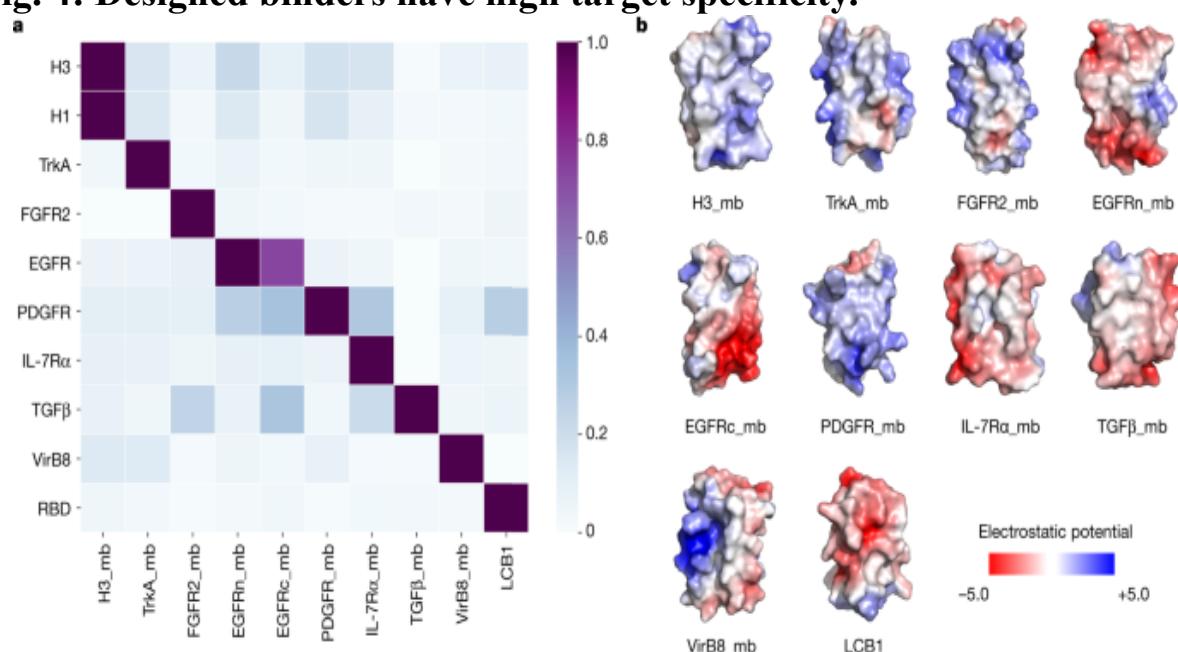
conformational rearrangements associated with membrane fusion, which is essential for virus infection^{32,33}. Neutralizing antibodies that target the stem region of group 2 HA have been identified through screens of large B cell libraries after vaccination or infection that neutralize both group 1 and group 2 influenza A viruses^{34,35}. Protein^{1,5}, peptide³⁶ and small-molecule inhibitors³⁷ have been designed to bind to the stem region of group 1 HA and neutralize influenza A viruses, but none recognize the group 2 HA. The design of small proteins or peptides that can bind and neutralize both group 1 and group 2 HA has been challenging owing to three main differences between group 1 and group 2 HA. First, the group 2 HA stem region is more hydrophilic, containing more polar residues. Second, in group 2 HA, Trp21 adopts a configuration roughly perpendicular to the surface of the targeting groove, which makes the targeted groove much shallower and less hydrophobic. And third, the group 2 HA is glycosylated at Asn38, with the carbohydrate side chains covering the hydrophobic groove (Extended Data Fig. 10a). We used our interface design method to design binders to H3 HA (A/Hong Kong/1/1968), the main pandemic subtype of group 2 influenza virus, and obtained a binder with an affinity of 320 nM to wild-type H3 (Fig. 2 and Table 1) and 28 nM to the deglycosylated H3 variant (N38D) (Extended Data Fig. 10b). The reduction in affinity is probably due to entropy loss of the glycan following binding and/or steric clash with the glycan. The binder also bound H1 HA (A/Puerto Rico/8/1934), which belongs to the main pandemic subtype of group 1 influenza virus (Extended Data Fig. 10b). The binding to both H1 and H3 HA is competed by the stem region that binds the neutralizing antibody FI6v3 (ref. 34) on the yeast surface (Extended Data Fig. 10c), which indicates that the binder attaches the HA at the targeted site. We also designed binders to the prokaryotic pathogen protein VirB8, part of the type IV secretion system of *R. typhi*, the causative agent of murine typhus²⁶. We selected the surface region composed of the second and the third helices of VirB8, and obtained binders with 510 pM affinity (Fig. 2 and Table 1).

With the outbreak of the SARS-CoV-2 pandemic, we applied our method to design miniproteins that targeted the receptor-binding domain of the SARS-CoV-2 spike protein near the ACE2 binding site to block receptor engagement. Owing to the pressing need for coronavirus therapeutics, we recently described the results of these efforts³⁸ ahead of those described in

this manuscript. Similar to FGFR2, IL-7R α and VirB8, the method yielded picomolar binders, which are among the most potent compounds known to inhibit the virus in cell culture (half-maximal inhibitory concentration (IC_{50}) of 0.15 ng ml^{-1}). Subsequent animal experiments showed that they provide potent protection against the virus *in vivo*³⁹. The modular nature of the miniprotein binders enables their rapid integration into designed diagnostic biosensors for both influenza and SARS-CoV-2 binders⁴⁰.

The designed binding proteins are all small proteins (<65 amino acids), and many are triple-helix bundles. To evaluate their target specificity, we tested the highest affinity binder to each target for binding to all other targets. There was little cross-reactivity (Fig. 4a), which is probably due to their diverse surface shapes and electrostatic properties (Fig. 4b). Consistent with previous observations with affibodies⁴¹, this result indicates that a wide variety of binding specificities can be encoded in simple helical bundles. In our approach, scaffolds are customized for each target, so the specificity arises both from the set of side chains at the binding interface and the overall shape of the interface itself.

Fig. 4: Designed binders have high target specificity.



To assess the cross-reactivity of each miniprotein binder (mb) with each target protein, biotinylated target proteins were loaded onto bilayer

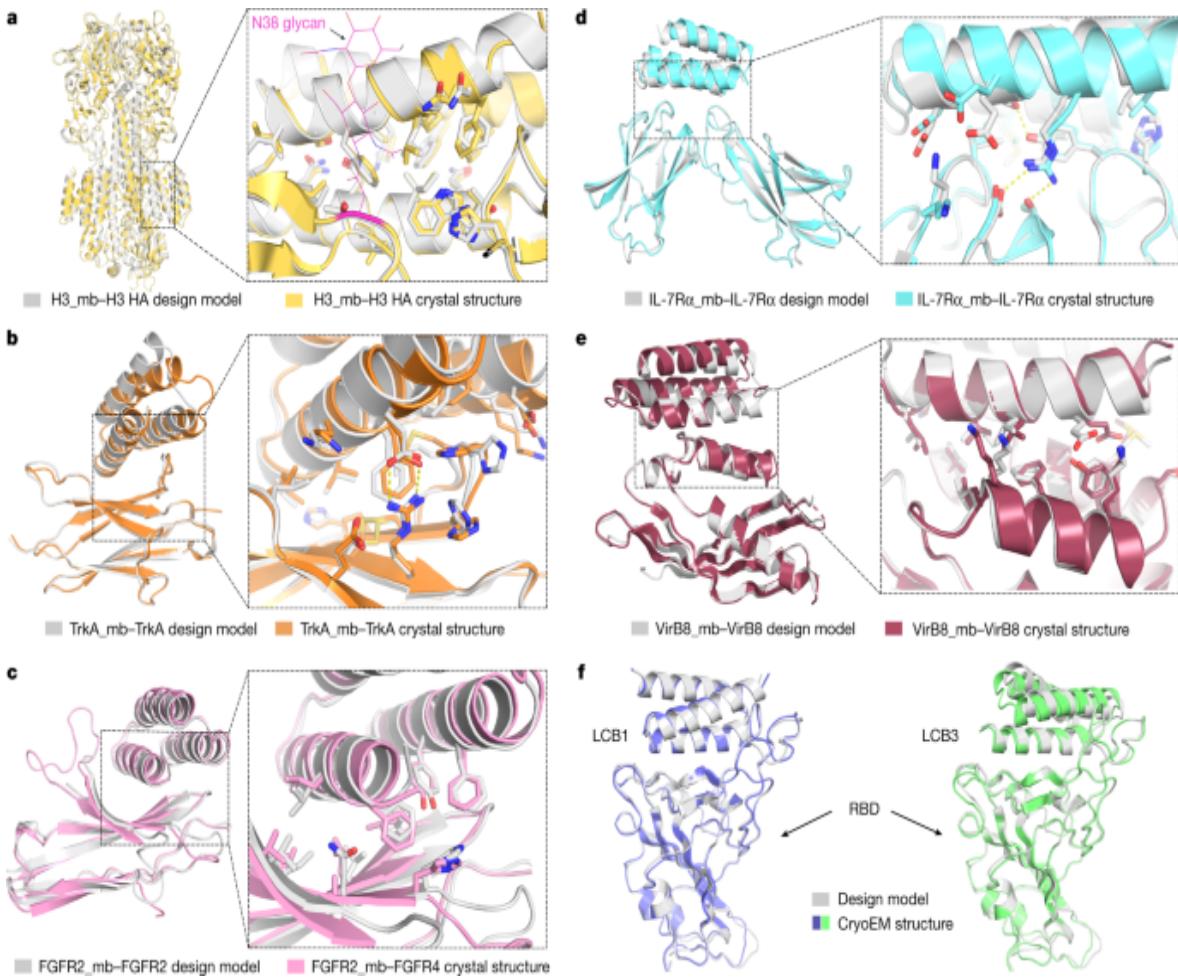
interferometry streptavidin sensors, allowed to equilibrate and the baseline signal set to zero. The biolayer interferometry tips were then placed into 100 nM binder solution for 5 min, washed with buffer, and dissociation was monitored for an additional 10 min. The heat map shows the maximum response signal for each binder–target pair normalized by the maximum response signal of the cognate designed binder–target pair. The raw biolayer interferometry traces are shown in the Supplementary Data 1. **b**, Surface shape and electrostatic potential (generated with the APBS Electrostatics plugin in PyMOL; red positive, blue, negative) of the designed binding interfaces.

High-resolution structural validation

High-resolution structures are crucial for evaluating the accuracy of computational protein designs. We succeeded in obtaining crystal structures of the unbound miniprotein binders for FGFR2 and IL-7R α , and co-crystal structures of the miniprotein binders of H3, TrkA, FGFR2, IL-7R α and VirB8 in complex with their targets (Extended Data Table 2).

The H3 binder bound to the shallow groove of the stem region of HK68/H3 HA in the crystal structure as designed. The C α root-mean-square deviation (r.m.s.d.) over the entire miniprotein binder was 1.91 Å using HA as the alignment reference (Fig. 5a). The binder makes extensive hydrophobic interactions with HA, and almost all of the designed interface side-chain configurations are recapitulated in the crystal structure (Fig. 5a). There was a clear reorientation of the oligosaccharide at Asn38 compared with the unbound HK68/H3 structure (Fig. 5a and Extended Data Fig. 10a; this has also been observed in HK68/H3 HA structures bound to stem region neutralizing antibodies^{34,35}). Consistent with this result, the binder has higher affinity for the N38D variant, which lacks this glycan, than for wild-type H3 HA (A/Hong Kong/1/1968) in biolayer interferometry assays (Table 1 and Extended Data Fig. 10b).

Fig. 5: High-resolution structures of miniprotein binders in complex with target proteins closely match the computational design models.



a–e, Left, superimposition of the computational design model (silver) on the experimentally determined crystal structure. Right, zoom-in view of the designed interface, with interacting side chains as sticks. **a**, H3 HA. **b**, TrkA. **c**, FGFR2. **d**, IL-7R α . **e**, VirB8. **f**, Superimposition of the computational design model and refined cryo-EM structures of LCB1 (left) and LCB3 (right) bound to the receptor-binding domain (RBD) of the SARS-CoV-2 spike protein.

The crystal structure of the TrkA binder in complex with TrkA was close to the design model (Fig. 5b). After aligning the crystal structure and design model on TrkA, the C_α r.m.s.d. over the entire miniprotein binder was 2.41 Å, and over the two interfacial binding helices, it was 1.20 Å. The crystal structures of the FGFR2 binder by itself (Extended Data Fig. 11a) and in complex with the third immunoglobulin-like domain of FGFR4 (Fig. 5c) matched the design models with near atomic accuracy, with C_α r.m.s.d.

values of 0.58 Å for the binder alone and 1.33 Å over the entire complex. The TrkA binder and the FGFR2 binder bound to the curved sheet side of the ligand-binding domain of TrkA and FGFR4, with extensive hydrophobic and polar interactions. Moreover, most of the key hydrophobic interactions as well as the primarily polar interactions in the computational design models were largely recapitulated in the crystal structures (Fig. 5b,c). The binding interfaces partially overlapped with the native ligand-binding sites of NGF and FGF; however, the side-chain interactions were entirely different in the designed and native complexes (Extended Data Fig. 3).

For IL-7R α , the crystal structure of the monomer was close to that of the design, with a C α r.m.s.d. of 0.63 Å (Extended Data Fig. 11b). The co-crystal structure with IL-7R α also closely matched that of the design model, with a C α r.m.s.d. of 2.2 Å using IL-7R α as the reference (Fig. 5d). Both the de novo IL-7R α binder and the native IL-7 use two helices to bind IL-7R α , but the binding orientations were different (Extended Data Fig. 3). The VirB8 binder made extensive interactions with the helical regions of VirB8 as designed; no native proteins have been identified to bind to this region. The C α r.m.s.d. over the entire miniprotein binder was 2.54 Å using the VirB8 as the alignment reference, and the side-chain configurations of key interface residues were largely recapitulated (Fig. 5e).

The heavy-atom r.m.s.d. values over the buried side chains at the interface (within 8 Å of the target in the design models) were 0.71 Å (H3), 1.10 Å (TrkA), 1.29 Å (FGFR2), 1.63 Å (IL-7R α) and 1.52 Å (VirB8), all of which are close to the core side-chain r.m.s.d. values (mean 0.90 Å). Further highlighting the accuracy of the protein interface design method, cryogenic electron microscopy (cryo-EM) structures of the SARS-CoV-2 binders LCB1 and LCB3 in complex with the virus were also nearly identical to the design models, with C α r.m.s.d. value of 1.27 Å and 1.9 Å, respectively³⁸ (Fig. 5f).

Although we were not able yet to solve structures for the remainder of the designs, the high-resolution sequence footprinting (Figs. 2 and 3) and competition results (Extended Data Fig. 7) suggest that the interfaces involve both the designed residues and the intended regions on the target. The close agreement between the experimentally determined structures and

the original design models indicates that the substitutions required to achieve high affinity play relatively subtle parts in tuning interface energetics: the overall structure of the complex, including the structure of the monomer binders and the detailed target binding mode, are determined by the computational design procedure.

Determinants of design success

For our de novo design strategy to be successful, we must encode in the approximately 60-residue designed sequences information on both the folded monomer structures and on the target binding interfaces. Indeed, designs that do not fold into the correct structure or that fold into the intended structures but do not bind to the target will fail. To assess the accuracy with which the monomer structure must be designed, we carried out an additional calculation and experiment for the IL-7R α target. Large numbers of scaffolds were superimposed onto 11 interface helical binding motifs identified in the first broad design search, and sequence design was carried out as described above. A strong correlation was found between the extent of binding and the root mean square deviations to the binding motif (Extended Data Fig. 12a), which indicates that designed backbones must be relatively accurate to achieve binding.

To assess the determinants of binding of the designed interfaces, assuming that the designs fold into the intended monomer structures, we took advantage of the large dataset (810,000 binder designs and 240,000 single mutants) generated in this study. Design success rates varied considerably between the different targets. For some (FGFR2 and PDGFR), hundreds of binders were generated, whereas for others (TIE2 and CD3 δ), fewer than 10 binders were obtained from libraries of 100,000 designs (Extended Data Table 1). Across all targets, there was a strong correlation between success rate and the hydrophobicity of the targeted region (Extended Data Fig. 12b), and designs observed experimentally to bind their targets tended to have stronger predicted binding energy and larger contact molecular surfaces (Extended Data Fig. 13). As found previously for designs of protein stability¹⁰, iterative design-build-test cycles in which the design method is updated at each iteration to incorporate feedback from the previous design

round should lead to systematic improvements in the design methodology and success rate.

Conclusions

Our success in designing nanomolar affinity binders for 14 target sites demonstrates that binding proteins can be designed *de novo* using only information on the structure of the target protein, without the need for prior information on binding hotspots or fragments from structures of complexes with binding partners. This success also suggests that our design pipeline provides a general solution to the *de novo* protein interface design problem that goes far beyond previously described methods. However, there is still considerable room for improvement. Only a small fraction of designs bind, and in almost all cases, the best of these require additional substitutions to achieve high-affinity binding. Furthermore, the design of binders to highly polar target sites remains a considerable challenge: the sites targeted here all contain at least four hydrophobic residues. The datasets generated in this work—both the information on binders versus non binders and the feedback on the effects of individual point mutants on binding—should help to guide the development of methods for designing high-affinity binders directly from the computer with no need for iterative experimental optimization. More generally, the *de novo* binder design method and the large dataset generated here provide a starting point to investigate the fundamental physical chemistry of protein–protein interactions and to develop and assess computational models of protein–protein interactions.

This work represents a major step forward towards the longer range goal of direct computational design of high-affinity binders starting from structural information alone. We anticipate that the binders created here, and new ones created with the method moving forwards, will find wide utility as signalling pathway antagonists as monomeric proteins and as tuneable agonists when rigidly scaffolded in multimeric formats, and in diagnostics and therapeutics for pathogenic disease. Unlike antibodies, the designed proteins are soluble when expressed in *E. coli* at high levels and are thermostable, and hence could form the basis for a next generation of lower cost protein therapeutics. More generally, the ability to rapidly and robustly design high-affinity

binders to arbitrary protein targets could transform the many areas of biotechnology and medicine that rely on affinity reagents.

Methods

Broad search stage

The crystal structures of HA (Protein Data Bank (PDB) identifier: [4FNK](#))²⁵, EGFR (PDB: [1MOX](#), [4UV7](#))^{17,42}, PDGFR (PDB: [3MJG](#))¹⁸, IR (PDB: [4ZXB](#))¹⁹, IGF1R (PDB: [5U8R](#))²⁰, TIE2 (PDB: [2GY7](#))²¹, IL-7R α (PDB: [3DI3](#))²², CD3 (PDB: [1XIW](#))²³, TGF β (PDB: [3KFD](#))²⁴ and VirB8 (PDB: [4O3V](#))²⁶ were refined in the Rosetta energy field constrained by experimental diffraction data. The crystal structures of TrkA (PDB: [1WWW](#))¹⁵ and FGFR2 (PDB: [1EV2](#))¹⁶ were refined with the Rosetta FastRelax protocol with coordinate constraints. The targeting chain or the selected targeting region were extracted and used as the starting point for docking and design. To run PatchDock¹¹, the scaffolds were mutated to poly-valine first, and default parameters were used to generate the raw docks. RifDock was used to generate the RIF by docking billions of individual disembodied amino acids to the selected targeting regions⁶. In detail, hydrophobic side-chain R-groups are docked against the target using a branch-and-bound search to quickly identify favourable interactions with the target, and polar side-chain R-groups are enumeratively sampled around every target hydrogen bond donor or acceptor. To identify backbone placements from which these interactions can be made, side-chain rotamer conformations are grown backwards for all R-group placements, and their backbone coordinates stored in a six-dimensional spatial hash table for rapid look-up. For the hierarchical searching protocol, the miniprotein scaffold library (50–65 residues in length) was docked into the field of the inverse rotamers using a branch-and-bound searching algorithm from low-resolution spatial grids to high-resolution spatial grids. For the PatchDock+RifDock protocols, the PatchDock outputs were used as seeds for the initial positioning of the scaffolds, and the docks were further refined in the finest resolution RIF. These docked conformations were further optimized to generate shape and chemically complementary interfaces using the Rosetta FastDesign protocol, activating between side-chain rotamer optimization and

gradient-descent-based energy minimization. Several improvements were added to the sequence design protocol to generate better sequences for both folding and binding. These included a better repulsive energy ramping strategy⁹, upweighting cross-interface energies, a pseudo-energy term penalizing buried unsatisfied polar atoms¹³ and a sequence profile constraint based on native protein fragments¹². Computational metrics of the final design models were calculated using Rosetta, which includes ddG, shape complementary and interface buried SASA, contact molecular surface, among others, for design selection. All the script and flag files to run the programs are provided in the [Supplementary Information](#).

Focused search stage

The binding energy and interface metrics for all the continuous secondary structure motifs (helix, strand and loop) were calculated for the designs generated in the broad search stage. The motifs with good interactions (based on binding energy and other interface metrics, such as SASA and contact molecular surface) with the target were extracted and aligned using the target structure as the reference. All the motifs were then clustered based on an energy based-TMalign-like clustering algorithm. In brief, all the motifs were sorted on the basis of the interaction energy with the target, and the lowest energy motif in the unclustered pool was selected as the centre of the first cluster. A similar score between this motif and every motif remaining in the unclustered pool was calculated based on the TMalign algorithm⁴³ without any further superimposition. Those motifs within a threshold similar score (default of 0.7) from the current cluster centre were removed from the unclustered pool and added to the new cluster. The lowest energy motif remaining in the unclustered pool was selected as the centre of the next cluster, and the second step was repeated. This process continued for subsequent clusters until no motifs remained in the unclustered pool. The best motif from each cluster was then selected based on the per-position weighted Rosetta binding energy, using the average energy across all the aligned motifs at each position as the weight. Around 2,000 best motifs were selected, and the scaffold library was superimposed onto these motifs using the MotifGraft mover⁴⁴. Interface sequences were future optimized, and computational metrics were computed for the final optimized designs as described in the broad search stage. CPU time requirements to produce

100,000 designed binders to be tested experimentally were typically around 100,000 CPU hours (usually at least ten times as many binders were computationally designed than were ordered).

Rapid Rosetta packing evaluation (the Predictor)

A severe speed mismatch existed between the docking methods (RifDock and focused search) and the subsequent full sequence design step. Although the docking methods can typically produce outputs every 1–3 s, the full sequence design can take upwards of 4 min. To remedy this situation, a step was designed to take about 20 s that would be more predictive than metrics evaluated on raw docks, but faster than the full sequence design.

A stripped down version of the Rosetta beta_nov16 score function was used to design only with hydrophobic amino acids. Specifically, fa_elec, lk_ball[iso,bridge,bridge_unclp], and the _intra_ terms were disabled as these proved to be the slowest energy methods by profiling. All that remained were Lennard–Jones, implicit solvation and backbone-dependent one-body energies (fa_dun, p_aa_pp, rama_pro). Additionally, flags were used to limit the number of rotamers built at each position ([Supplementary Information](#)).

After the rapid design step, the designs were minimized twice: once with a low-repulsive score function and again with a normal-repulsive score function. Metrics of interest were then evaluated, including like Rosetta ddG, contact molecular surface, and contact molecular surface to critical hydrophobic residues.

Based on the observation that these predicted metrics correlated with the values after full sequence design, a maximum likelihood estimator (a functional form similar to logistic regression) was used to give each predicted design a likelihood that it should be selected to move forward. A subset of the docks to be evaluated were subjected to the full sequence design, and their final metric values calculated. With a goal threshold for each filter, each fully designed output can be marked as pass or fail for each metric independently. Then, by binning the fully designed outputs by their values from the rapid trajectory and plotting the fraction of designs that pass the goal threshold, the probability that each predicted design passes each

filter can be calculated (sigmoids are fitted to smooth the distribution). From here, the probability of passing each filter may be multiplied together to arrive at the final probability of passing all filters. This final probability can then be used to rank the designs and pick the best designs to move forward to full sequence optimization.

Note that the rapid design protocol here is used merely to rank the designs, not to optimize them; the raw, non-rapid-designed docks are the structures carried forward.

Contact molecular surface

SASA is a measure of the exposure of amino acids to the solvent and it is typically calculated using methods that involve in silico rolling of a spherical probe, which approximates a water molecule (radius 1.4 Å), around a full-atom protein model. Delta-SASA after protein–protein binding has been widely used to analyse native protein interactions. Unlike the crystal structures of the native protein complexes, design models for the de novo interactions are usually imperfectly packed and contain many holes or cavities. If the sizes of the holes or cavities in the interface are smaller than the rolling probe, SASA cannot capture those holes and cavities and the real contacts are usually overestimated by the delta-SASA metric. The contact molecular surface was developed to mitigate the flaws of the de novo designed interactions. First, the molecule surfaces of the binder and the target were calculated using the triangularization algorithm in the Rosetta shape complementary filter. For each triangle, the distance to the closest triangle on the other side was calculated and used to downweight the area of the triangle by the following equation: $A' = A \times \exp(-0.5 \times \text{distance}^2)$. Then all the downweighted areas were summed to obtain the contact molecular surface. In this way, the real contacts between the target and the binder are penalized by the cavities and holes in the interface. The contact molecular surface was implemented as the ContactMolecularSurface filter in the Rosetta macromolecular modelling suite.

Upweighted protein interface interactions

Rosetta sequence design starts from generating an interaction graph by calculating the energies between all designable rotamer pairs⁴⁵. The best rotamer combinations are searched using a Monte Carlo simulated annealing protocol by optimizing the total energy of the protein (monomer/complex). To obtain more contacts between the binder and the target protein, we can upweight the energies of all the cross-interface rotamer pairs by a defined factor. In this way, the Monte Carlo protocol will be biased to find solutions with better cross-interface interactions. The upweighted protein interface interaction protocol was implemented as the ProteinProteinInterfaceUpweighter task operation in the Rosetta macromolecular modelling suite.

Comparison of sampling efficiency of PatchDock, RifDock and resampling protocols

The top 30 PatchDock outputs for the 1,000 helical scaffolds tested were designed using the RosettaScripts protocol. RifDock seeded with PatchDock outputs generated 300 outputs per scaffold, which were trimmed to a total of 19,500 docks with the Predictor ([Methods](#)) and subsequently designed. The top 150 RifDock outputs per scaffold were trimmed to 9,750, designed, and 300 motifs were extracted. The motifs were grafted into the scaffold set to produce 150,000 docks, which were trimmed to 9,750, designed, and combined with the earlier 9,750.

DNA library preparation

All protein sequences were padded to 65 amino acids by adding a (GGGS) n linker at the carboxy terminus of the designs to avoid the biased amplification of short DNA fragments during PCR reactions. The protein sequences were reversed translated and optimized using DNAworks2.0 (ref. [46](#)) with the *Saccharomyces cerevisiae* codon frequency table.

Oligonucleotide pools encoding the de novo designs and the point mutant library were purchased from Agilent Technologies. Combinatorial libraries were purchased as Integrated DNA Technologies ultramers, with the final DNA diversity ranging from 1×10^6 to 1×10^7 .

All libraries were amplified using Kapa HiFi polymerase (Kapa Biosystems) with a qPCR machine (Bio-Rad, CFX96). In detail, the libraries were first amplified in a 25 µl reaction, and the PCR reaction was terminated when the reaction reached half maximum yield to avoid overamplification. The PCR product was loaded onto a DNA agarose gel. The band with the expected size was cut out, and DNA fragments were extracted using QIAquick kits (Qiagen). Then, the DNA product was re-amplified as before to generate enough DNA for yeast transformation. The final PCR product was cleaned up with a QIAquick Clean up kit (Qiagen). For the yeast transformation step, 2–3 µg of linearized modified pETcon vector (pETcon3) and 6 µg of insert were transformed into the EBY100 yeast strain using a previously described protocol^{[47](#)}.

DNA libraries for deep sequencing were prepared using the same PCR protocol, except the first step started from yeast plasmid prepared from 5×10^7 to 1×10^8 cells by Zymoprep (Zymo Research). Illumina adapters and 6-bp pool-specific barcodes were added in the second qPCR step. Gel extraction was used to obtain the final DNA product for sequencing. All the different sorting pools were sequenced using Illumina NextSeq sequencing.

Target protein preparation

The influenza A HA ectodomain was expressed using a baculovirus expression system as previously described^{[25,48](#)}. In brief, each HA was fused with a gp67 signal peptide at the amino terminus and to a BirA biotinylation site, thrombin cleavage site, trimerization domain and His-tag at the C terminus. Expressed HA was purified using metal affinity chromatography with Ni²⁺-NTA resin. For binding studies, each HA was biotinylated with BirA and purified by gel filtration using a S200 16/90 column on an ÄKTA protein purification system (GE Healthcare). The biotinylation reactions contained 100 mM Tris (pH 8.5), 10 mM magnesium acetate, 10 mM ATP, 50 µM biotin and <50 mM NaCl, and were incubated at 37 °C for 1 h.

For TrkA, the DNA encoding the human TrkA extracellular domain (ECD) (residues 36–382) was cloned into pAcBAP, a derivative of pAcGP67-A modified to include a C-terminal biotin acceptor peptide (BAP) tag (GLNDIFEAQKIEWHE) followed by a 6×His tag for affinity purification.

It was then transfected into *Trichoplusia ni* (High Five) cells (Invitrogen) using the BaculoGold baculovirus expression system (BD Biosciences) for secretion and purified from the clarified supernatant through Ni-NTA followed by size-exclusion chromatography (SEC) with a Superdex-200 column in sterile PBS (Gibco, 20012-027). The ectodomains of FGFR2 (residues 147–366, UniProt ID: [P21802](#)), EGFR (residues ID 25–525, UniProt ID: [P00533](#)), PDGFR (residues 33–314, UniProt ID: [P09619](#)), IR (residues ID 28–953, UniProt ID: [P06213](#)), IGF1R (residues 31–930, UniProt ID: [P08069](#)), TIE2 (residues 23–445, UniProt ID: [Q02763](#)), IL-7R α (residues 37–231, UniProt ID: [P16871](#)) were expressed in mammalian cells with a IgK signal peptide (METDTLLLWVLLLWVPGSTG) at the N terminus and a C-terminal tag (GSENLKYFQGSHHHHHGSGLNDIFEAQKIEWHE) that contains a TEV cleavage site, a 6-His tag and an AviTag. VirB8 was expressed in *E. coli* with a C-terminal AviTag as previously described²⁶. The proteins were purified by Ni²⁺-NTA, and polished by SEC. The AviTag proteins were then biotinylated with a BirA biotin-protein ligase bulk reaction kit (Avidity) following the manufacturer's protocol, and the excess biotin was removed through SEC. Biotinylated CD3 protein was purchased from Abcam (ab205994). TGF β was purchased from Acro Biosystems (TG1-H8217). IGF1 was purchased from Sigma (407251-100 μ g). Insulin was purchased from Abcam (ab123768). The caged ANG1-Fc protein was prepared as previously described⁴⁹, and was provided by G. Ueda. The FI6v3 antibody was provided by D. H. Fuller (University of Washington).

Yeast surface display

Saccharomyces cerevisiae EBY100 strain cultures were grown in C-Trp-Ura medium supplemented with 2% (w/v) glucose. For induction of expression, yeast cells were centrifuged at 6,000g for 1 min and resuspended in SGCAA medium supplemented with 0.2% (w/v) glucose at the cell density of 1×10^7 cells per ml and induced at 30 °C for 16–24 h. Cells were washed with PBSF (PBS with 1% (w/v) BSA) and labelled with biotinylated targets using two labelling methods: with-avidity and without-avidity labelling. For the with-avidity method, the cells were incubated with biotinylated target, together with anti-c-Myc fluorescein isothiocyanate (FITC, Miltenyi Biotech) and streptavidin–phycoerythrin (SAPE, ThermoFisher). The concentration of

SAPE in the with-avidity method was used at one-quarter of the concentration of the biotinylated targets. For the without-avidity method, the cells were first incubated with biotinylated targets, washed and secondarily labelled with SAPE and FITC. All the original libraries of de novo designs were sorted using the with-avidity method for the first few rounds of screening to exclude weak binder candidates, followed by several without-avidity sorts with different concentrations of targets. For SSM libraries, two rounds of without-avidity sorts were applied and in the third round of screening, the libraries were titrated with a series of decreasing concentrations of targets to enrich mutants with beneficial mutations. The combinatorial libraries were sorted to convergence by decreasing the target concentration with each subsequent sort and collecting only the top 0.1% of the binding population. The final sorting pools of the combinatorial libraries were plated on C-trp-ura plates, and the sequences of individual clones were determined by Sanger sequencing. The competition sort was done following the without-avidity protocols with a minor modification. In brief, the biotinylated target proteins (H1, H3, TrkA, IR, IGF1R, PDGFR and TIE2) were first incubated with an excess amount of competitors (FI6v3, FI6v3, NGF, insulin, IGF1, PDGF and caged ANG1-Fc, respectively) for 10 min, and the mixture was used for labelling the cells. The nonspecificity reagent was prepared using the protocol as previously described⁵⁰. For the nonspecificity sort, the cells were first washed with PBSF and incubated with the nonspecificity reagent at a concentration of 100 µg ml⁻¹ for 30 min. The cells were then washed and secondarily labelled with SAPE and FITC for cell sorting. The cells were then labelled with RBD using the above-described protocol.

Miniprotein expression

Genes encoding the designed protein sequences were synthesized and cloned into modified pET-29b(+) *E. coli* plasmid expression vectors (GenScript, N-terminal 8-His tag followed by a TEV cleavage site). For all the designed proteins, the sequence of the N-terminal tag is MSHHHHHHHSENLYFQSGGG (unless otherwise noted), which is followed immediately by the sequence of the designed protein. For proteins expressed with the maltose binding protein (MBP) tag, the corresponding genes were subcloned into a modified pET-29b(+) *E. coli* plasmid, which

has a N-terminal 6-His tag and a MBP tag. Plasmids were transformed into chemically competent *E. coli* Lemo21 cells (NEB). For the designs for TrkA, FGFR2, EGFR, IR, IGF1R, TIE2, IL-7R α , TGF β and the MBP-tagged miniproteins, protein expression was performed using Studier autoinduction medium supplemented with antibiotic, and cultures were grown overnight. For the HA, PDGFR and CD3 δ designs, the *E. coli* cells were grown in LB medium at 37 °C until the cell density reached 0.6 at OD₆₀₀. Then, IPTG was added to a final concentration of 500 mM and the cells were grown overnight at 22 °C for expression. The cells were collected by spinning at 4,000g for 10 min and then resuspended in lysis buffer (300 mM NaCl, 30 mM Tris-HCL (pH 8.0), with 0.25% CHAPS for cell assay samples) with DNase and protease inhibitor tablets. The cells were lysed with a Qsonica Sonicators sonicator for 4 min in total (2 min each time, 10 s on, 10 s off) with an amplitude of 80%. The soluble fraction was clarified by centrifugation at 20,000g for 30 min. The soluble fraction was purified by immobilized metal affinity chromatography (Qiagen) followed by FPLC SEC (Superdex 75 10/300 GL, GE Healthcare). All protein samples were characterized by SDS-PAGE, and purity was greater than 95%. Protein concentrations were determined by absorbance at 280 nm measured with a NanoDrop spectrophotometer (Thermo Scientific) using predicted extinction coefficients.

Circular dichroism

Far-ultraviolet circular dichroism measurements were carried out with a JASCO-1500 instrument equipped with a temperature-controlled multi-cell holder. Wavelength scans were measured from 260 to 190 nm at 25 and 95 °C and again at 25 °C after fast refolding (about 5 min). Temperature melts monitored the dichroism signal at 222 nm in steps of 2 °C min⁻¹ with 30 s of equilibration time. Wavelength scans and temperature melts were performed using 0.3 mg ml⁻¹ protein in PBS buffer (20 mM NaPO₄, 150 mM NaCl, pH 7.4) with a 1 mm path-length cuvette. Melting temperatures were determined by fitting the data with a sigmoid curve equation. Nine out of the 13 designs retained more than half of the mean residue ellipticity values, which indicated that the T_m values are greater than 95 °C. T_m values of the other designs were determined as the inflection point of the fitted function.

Biolayer interferometry

Biolayer interferometry binding data were collected on an Octet RED96 (ForteBio) and processed using the instrument's integrated software. For minibinder binding assays, biotinylated targets were loaded onto streptavidin-coated biosensors (ForteBio) at 50 nM in binding buffer (10 mM HEPES (pH 7.4), 150 mM NaCl, 3 mM EDTA, 0.05% surfactant P20 and 1% BSA) for 6 min. Analyte proteins were diluted from concentrated stocks into the binding buffer. After baseline measurement in the binding buffer alone, the binding kinetics were monitored by dipping the biosensors in wells containing the target protein at the indicated concentration (association step) and then dipping the sensors back into baseline/buffer (dissociation). The binding affinities of TIE2 and IGF1R minibinders were low, and MBP-tagged proteins were used for the binding assay to amplify the binding signal. The binding assay for the IR designs were conducted with Amine Reactive Second-Generation (AR2G ForteBio) Biosensors with the recommended protocol. In brief, the miniproteins were immobilized onto the AR2G tips and the IR sample was used as the analyte with the indicated concentrations. Data were analysed and processed using ForteBio Data Analysis software v.9.0.0.14.

For the cross-reactivity assay, each target protein was loaded onto streptavidin tips at a concentration of 50 nM for 325 s. The tips were dipped into the miniprotein wells for 300 s (association) and then dipped into the blank buffer wells for 600 s (dissociation). The maximum raw biolayer interferometry signal binding was used as the indicator of binding strength. The maximum signal among all the miniprotein binders for a specific target was used to normalize the data for heat-map plotting.

Crystallization and structure determination of the H3 binder in complex with HK68/H3

To prepare the H3 minibinder (H3_mb)-HK68/H3 HA complex for crystallization, a fivefold molar excess of H3_mb was mixed with about 2 mg ml⁻¹ of HK68/H3 HA in 20 mM Tris (pH 8.0), 150 mM NaCl. The mixture was incubated overnight at 4 °C to facilitate formation of the complex. Saturated complexes were then purified from unbound HB_mb by

gel filtration. Gel filtration fractions containing the H3[−]mb–HK68/H3 HA complex were concentrated to approximately 7 mg ml^{−1} in 20 mM Tris (pH 8.0) and 150 mM NaCl. Crystallization screens were set up using the sitting-drop vapour-diffusion method with our automated CrystalMation robotic system (Rigaku) at The Scripps Research Institute. Within 3–7 days, diffraction-quality crystals had grown in 0.2 M sodium thiocyanate and 20% (w/v) PEG 3350 as a precipitant. The resulting crystals were cryoprotected through the addition of 5–15% ethylene glycol, flash cooled and stored in liquid nitrogen until data collection. Diffraction data were collected at 100 K at the Stanford Synchrotron Radiation Lightsource (SSRL) beamline 12-1 and processed with HKL-2000 (ref. [51](#)). Initial phases were determined by molecular replacement using Phaser^{[52,53](#)} with a HA model from PDB identifier [4FNK](#) (apo HK68/H3 HA). Refinement was carried out in Phenix^{[54](#)}, alternating with manual rebuilding and adjustment in COOT^{[55](#)}. Electron-density maps were calculated using Phenix Data collection, and refinement statistics are summarized in Extended Data Table [2](#). The final coordinates were validated using MolProbity^{[56](#)}.

Crystal structure of TrkA in complex with the miniprotein binder

The human TrkA receptor ECD was produced in insect cells using baculovirus and prepared as previously described^{[57](#)}. Hi5 cells were co-infected in shaking Fernbach flasks with baculoviruses encoding TrkA ECD and endoglycosidase H in the presence of kifunensine. Cultures were allowed to progress for 65 h before the supernatant was recovered by centrifugation. Components from the medium were precipitated by the addition of 50 mM Tris (pH 8.0), 1 mM NiCl₂ and 5 mM CaCl₂, and the supernatant was filtered over diatomaceous earth. The filtrate was batch-bound to Ni²⁺-NTA resin, eluted with 200 mM imidazole in HBS (HEPES-buffered saline: 10 mM HEPES (pH 7.3), 150 mM NaCl), and purified by SEC on a Superdex-75 column (Cytiva Life Sciences). To prepare the TrkA–miniprotein complex, an excess amount of miniprotein was mixed with TrkA, digested overnight at 4 °C with 1:100 (w/w) carboxypeptidases A and B, and purified by SEC.

For crystallization, the TrkA–ligand complex was concentrated to 38 mg ml⁻¹ in HBS and screened in sitting-drop format using a Mosquito crystallization robot (SPT Labtech). Initial sea urchin-like crystals were obtained from the MCSG1 screen (Anatrace-Microlytic) in 0.17 M ammonium acetate, 0.085 M sodium citrate (pH 5.6), 25.5% PEG 4000 and 15% glycerol. These crystals were crushed and used to microseed the MCSG1 screen again at a ratio of 3:2:1 protein:precipitant:seed stock, resulting in single plate-like crystals grown from 0.2 M ammonium sulfate, 0.1 M bis-Tris (pH 6.5) and 25% PEG 3350. After further optimization to 0.4 M ammonium sulfate, 0.1 M bis-Tris (pH 6.2) and 20% PEG 3350, new seeds were prepared for final seeding into 0.4 M ammonium sulfate, 0.1 M bis-Tris (pH 6.2) and 16% PEG 3350.

Crystals were cryoprotected by the addition of ethylene glycol to 30% (v/v) and flash cooled in liquid nitrogen. Diffraction data to 1.84 Å resolution were collected at 100 K using an X-ray wavelength of 1.033 Å at the SSRL beamline 12-2. Crystals were assigned to space group P21 with unit cell dimensions $a = 42.20 \text{ \AA}$, $b = 205.70 \text{ \AA}$, $c = 72.57 \text{ \AA}$ and $\beta = 106.42^\circ$. Data were indexed, integrated and scaled using XDS^{58,59} and merged using Pointless and Aimless from the CCP4 suite^{60,61,62}.

The structure was solved by molecular replacement in Phaser⁵² using separated domains of TrkA ECD (PDB accession [2IFG](#)) and the predicted model of the ligand as search models to place two copies of the complex in the asymmetric unit. Initial rebuilding was completed with phenix.autobuild⁶³ followed by iterative rounds of manual rebuilding in Coot⁶⁴ and refinement in Phenix^{65,66,67}. TLS parameters were chosen using TLSMD⁶⁸, and NCS restraints were used throughout refinement⁶⁹. The final resolution of the data was selected as 1.84 Å by comparing the results of paired refinements at 1.84, 1.90, 1.95, 2.00 and 2.05 Å resolution⁷⁰. The final refined model included 97.26% of residues in the favoured region of the Ramachandran plot with 0.25% outliers as calculated by MolProbity⁵⁶.

Crystallographic software used in this study was configured and installed by SBGrid⁷¹. Diffraction images have been deposited in the SBGrid Data Bank with the identifier [839](#), and the final model and reflections have been deposited in the PDB with the identifier [7N3T](#).

Crystal structures of FGFR2_mb in complex with FGFR4 domain 3 and FGFR2_mb alone

cDNA of human FGFR4 domain 3 (FGFR4_{D3}, amino acids S245–D355) was amplified by PCR and cloned into pET-28a(+) plasmid (Novagen). The plasmid containing FGFR4_{D3} with N-terminal hexa-histidine tag was transformed into BL21(DE3) cells. The transformed cells were grown in LB medium at 37 °C until the OD₆₀₀ reached 0.5, induced with 1.0 mM IPTG, grown for an additional 4 h at 37 °C and collected. The bacterial cells were resuspended and lysed by sonication. FGFR4_{D3} was refolded from insoluble fractions using a previously reported procedure^{16,72,73}, and purified to homogeneity using nickel affinity chromatography (Ni²⁺-NTA agarose; Qiagen) followed by SEC (Superdex 200 Increase 10/300 GL, Cytiva) equilibrated with a buffer containing 200 mM NaCl, 25 mM HEPES (pH 8.0) and 5% glycerol. The purified FGFR4_{D3} was mixed with a 1.2-fold molar excess of FGFR2_mb and subjected to another round of SEC to isolate the FGFR4_{D3}–FGFR2_mb complex. Fractions containing FGFR4_{D3} bound to FGFR2_mb were pooled and concentrated to 12 mg ml⁻¹ and screened for crystallization using commercially available crystallization screening kits with Mosquito Crystal liquid handler (SPT Labtech). Crystals of the FGFR4_{D3}–FGFR2_mb complex were obtained with ProPlex screening solution (Molecular Dimensions) containing 0.2 M sodium chloride, 0.1 M MES pH 6.0 and 20% PEG 3,350 at 4 °C. The crystals were cryoprotected using the mother liquor supplemented with 25% glycerol before being flash-cooled in liquid nitrogen.

Crystals of FGFR2_mb were obtained using solution containing alcohols (0.02 M 1,6-hexanediol, 0.02 M 1-butanol, 0.02 M 1,2-propanediol, 0.02 M 2-propanol, 0.02 M 1,4-butanediol, 0.02 M 1,3-propanediol), buffer mixture (0.1 M Tris and BICINE adjusted to pH 8.5) and precipitants (12.5% v/v MPD, 12.5% PEG 1000, 12.5% w/v PEG 3,350) by the hanging-drop vapour-diffusion method at 20 °C, which were directly flash-cooled in liquid nitrogen for X-ray diffraction data collection.

X-ray diffraction data were collected at the NE-CAT 24ID-E beam line of Advanced Photon Source (Argonne National Laboratory) and processed

with XDS⁷⁴. The initial structure of FGFR2_mb was obtained by molecular replacement with PHASER^{52,75} using the designed model, which was iteratively refined using PHENIX^{67,75} followed by manual building with COOT⁶⁴. The structure of FGFR4_{D3}–FGFR2_mb complex was obtained by molecular replacement with Phaser^{52,75} using the coordinates corresponding to the domain 3 region of FGFR1c⁷² (PDB ID: [1CVS](#)) and the coordinates of FGFR2_mb as the search model, followed by iterative refinements using PHENIX^{67,75} and COOT⁶⁴. The final structures were validated with MolProbity^{75,76}. Data collection and refinement statistics are provided in Extended Data Table [2](#).

Crystal structure of unbound IL-7Ra minibinder

To facilitate crystallization, the N-terminal His-tag was removed using TEV protease and the protein was concentrated to 40 mg ml⁻¹ in 30 mM Tris-HCl (pH 8.0) and 150 mM NaCl. Sparse-matrix crystal screening was performed using kits from Hampton Research (Index-HT, PEGRx-HT and PEG/Ion-HT) at room temperature. A Mosquito nanolitre crystallization robot was used to set up sitting drops consisting of 200 nl of protein and 200 nl of each reservoir solution with 80 µl of reservoir solution in MRC-2 plates.

Promising prism-shaped crystals grew from the IndexHT C3 condition, and optimal conditions ranged from 2.4 to 3.0 M sodium malonate (pH 7.0). Protein crystals were cryo-cooled directly into liquid nitrogen. Initial X-ray diffraction experiments were carried out on a home-source system equipped with MicroMax-007 HF rotating anode with a Dectris Eiger R 4M single-photon counting device. X-ray diffraction data on optimized protein crystals were collected at the Advanced Photon Source synchrotron beamline 23ID-D of GM/CA with a Dectris Pilatus3-6M detector. All X-ray data were processed with XDS. Molecular replacement using the de novo designed model was used to solve the crystal structure using Phaser within the Phenix package. Two molecules were located in the asymmetric unit. Structural refinement used Phenix using no NCS restraints. Data collection and refinement statistics are given in Extended Data Table [2](#).

Crystal structure of IL-7Ra in complex with the minibinder

The ectodomain of human IL-7R α was produced and purified as previously described⁷⁷. The anti-IL-7R α minibinder was prepared as described above. The IL-7R α -minibinder complex was formed by adding a molar excess of purified minibinder to recombinant IL-7R α . The IL-7R α -minibinder complex was purified by SEC using a Superdex-75 column (Cytiva Life Sciences) with HBS buffer (pH 7.4) as the running buffer. Fractions corresponding to the IL-7R α -minibinder complex were pooled and concentrated by centrifugal ultrafiltration to a concentration of 3.9 mg ml⁻¹. Sparse-matrix crystallization screens were carried out using the BCS-Screen (Molecular Dimensions) at 293 K and the sitting-drop method. The vapour-diffusion geometry was used to set up sitting drops consisting of 200 nl of protein and 100 nl of each reservoir solution using a Mosquito nanolitre crystallization robot (TTP Labtech). The IL-7R α -minibinder complex crystallized in condition A5 (0.1 M phosphate, citrate (pH 5.5) and 25.0% PEG Smear medium). Crystals were cryo-protected with mother liquor supplemented with 25% v/v PEG 400 and cryo-cooled by direct plunging into liquid nitrogen. X-ray diffraction data of protein crystals were collected at beamline ID23-2 of the ESRF (Grenoble) with a Dectris PILATUS3 X 2M detector and were processed with XDS⁵⁸. The structure was determined by maximum-likelihood molecular replacement in Phaser using the crystal structure of IL-7R α (PDB ID: [3DI2](#)) as a search model⁵². Three copies of the complex were located in the asymmetric unit. Model (re)building was performed in Coot⁶⁴, and coordinate and ADP refinement was performed in PHENIX⁶⁵ and autoBuster⁷⁸. Model and map validation tools in Coot, the PHENIX suite and the PDB_REDQ server⁷⁹ were used to validate the quality of crystallographic models. The final model and reflections have been deposited in PDB with the identifier [7OPB](#). Data collection and refinement statistics are provided in Extended Data Table [2](#).

Crystal structure of VirB8-like protein in complex with the minibinder

VirB8-like protein of the type IV secretion system from *R. typhi* (UniProt ID: [Q68X84](#)) in complex with 0.75 mM VirB8 miniprotein binder was suspended in a buffer containing 20 mM HEPES pH 7.0, 300 mM NaCl and 5% glycerol. The complex was crystallized using the sitting-drop vapour-diffusion method at 14 °C with drops composed of 0.4 ml of the complex at

9.9 mg ml⁻¹ mixed with 0.4 ml crystallant (sparse matrix screen JCSG Top96 (Rigaku Reagents) condition G9: 100 mM sodium acetate/hydrochloric acid (pH 4.6), 25% (w/v) PEG 4000, 200 mM ammonium sulfate) equilibrated against 80 ml crystallant in the reservoir. Crystals were cryoprotected in the crystallant supplemented with 15% (v/v) ethylene glycol. X-ray diffraction data of the VirB8 protein–miniprotein binder complex was collected at the LS-CAT beamline 21-ID-F at the Advanced Photon Source. Data were integrated in XDS and reduced using XSCALE⁵⁸. Data quality was assessed using POINTLESS⁸⁰. Molecular replacement was performed using Phaser⁵² with search models comprising a previously solved crystal structure of *R. typhi* VirB8-like of type IV secretion system (PDB ID: [4O3V](#)) and an Alphafold2 (ref. [81](#)) predicted model of the VirB8 miniprotein binder. Iterative manual model building and refinement were carried out using Coot⁶⁴ and Phenix⁶⁵. Structure quality was assessed using Molprobity⁵⁶ before deposition in the PDB^{[82,83](#)} (Extended Data Table [2](#)). Diffraction images are available at the Integrated Resource for Reproducibility in Macromolecular Crystallography^{[84,85](#)}.

Comparison between the crystal structures and design models

For the structures of the miniprotein binders in complex with the targets, the entire structures were aligned using the target as the references first. The r.m.s.d. over the C_α atoms of the entire miniprotein binder was calculated. For the unbound crystal structures of the FGFR2 miniprotein binder and the IL-7Ra miniprotein binder, the r.m.s.d. values were calculated over all the C_α atoms after superimposition. For the analysis of the heavy atoms of the interface core residues, the structures were aligned using the target as references first. Interface residues of the binders were selected as long as there is one residue on the target that has a Cβ–Cβ distance of less than 8 Å using the NeighborhoodResidueSelector, and core residues were selected using the LayerSelector in Rosetta with the default burial cut-off value. Then heavy atoms of the interface core residues were used to calculate the r.m.s.d. values. Four, eight, six and six residues were considered as interface core residues for the H3, FGFR2, IL-7Ra and VirB8 complex structures respectively.

TrkA minibinder antagonist assay

The Phospho-flow signalling assay was used to characterize the antagonistic properties of the TrkA minibinder. TF-1 cells (American Type Culture Collection, CRL-2003) were starved for 4 h in base medium without NGF or other cytokines before signalling assays. Cells were plated in 96-well plates with different concentrations of TrkA binder and stimulated with human beta-NGF (R&D) for 10 min at 37 °C, followed by fixation with 1.6% paraformaldehyde for 10 min at room temperature. Cells were permeabilized by resuspension in ice-cold methanol and stored at –20 °C until flow cytometry analysis. For intracellular staining, the permeabilized cells were washed and incubated with Alexa Fluor-488 conjugated anti-ERK1/2 pT202/pY204 antibody (BD) and Alexa Fluor-647 conjugated anti-Akt pS473 antibody (Cell Signaling Technology) for 1 h at room temperature. After washing with autoMACS running buffer (Miltenyi), the fluorescence intensity of each antibody staining level was acquired using a CytoFlex flow cytometer (Beckman Coulter). Mean fluorescence intensity (MFI) values were background subtracted and normalized to the maximal MFI value in the absence of TrkA binder and plotted in Prism 9 (GraphPad). The dose–response curves were generated using the sigmoidal dose–response analysis method.

For the cell proliferation assay, TF-1 cells were plated in a 96-well plate and cultured in RPMI-1640 medium containing 2% FBS and different concentrations of TrkA binder and NGF for 48 h at 37 °C. The cell proliferation rate was assessed by measuring the cellular ATP level using CellTiter-Glo 2.0 Cell Viability Assay reagent (Promega) according to the manufacturer’s protocol. The luminescent signal was measured using a SpectraMax Paradigm plate reader, and the data were plotted and analysed using Prism 9 (GraphPad). The dose–response curves were generated using the sigmoidal dose-response analysis method.

FGFR2 and EGFR minibinder antagonist assay

For cell culture, human umbilical vein endothelial cells (HUVECs; Lonza, C2519AS) were grown in EGM2 medium on 35-mm cell culture dishes coated with 0.1% gelatin. In brief, EGM2 is composed of 20% FBS, 1%

penicillin–streptomycin, 1% GlutaMAX (Gibco, 35050061), 1% ECGS (endothelial cell growth factor), 1 mM sodium pyruvate, 7.5 mM HEPES, 0.08 mg ml⁻¹ heparin and 0.01% amphotericin B in a mixture of 1× RPMI-1640 with and without glucose (final glucose concentration = 5.6 mM). Medium was filtered through a 0.2-μm filter. HUVECs were serially passaged and expanded before cryopreservation.

FGFR and EGFR antagonist assay

Frozen HUVECs were thawed and cultured in a 35-mm dish in EGM2 medium until confluence was reached. After that, EGM2 medium was aspirated and cells were rinsed twice with 1× PBS. Cells were then serum-starved by adding 2 ml of DMEM serum-free medium (1 g l⁻¹ glucose, Gibco) for 16 h, after which the starvation medium was aspirated. The cells were then treated with the FGFR2 minibinder or the EGFR minibinder for 1 h at 37 °C and at concentrations varying between 5 nM and 1 μM of minibinder. This was followed by stimulation with β-FGF (0.75 nM, Fisher Scientific) or EGF (1 nM, Peprotech), respectively, for 15 min at 37 °C. After treatment, the medium was aspirated, and cells were washed once with 1× PBS before collecting the total protein for analysis.

Total protein isolation

After minibinder treatment, the cells were gently rinsed in 1× PBS before lysis with 130 μl of lysis buffer containing 20 mM Tris-HCL (pH 7.5), 150 mM NaCl, 15% glycerol, 1% Triton, 3% SDS, 25 mM β-glycerophosphate, 50 mM NaF, 10 mM sodium pyrophosphate, 0.5% orthovanadate, 1% PMSF (all obtained from Sigma-Aldrich), benzonase nuclease (EMD Chemicals), protease inhibitor cocktail (Pierce protease inhibitor mini tablets, Thermo Scientific) and phosphatase inhibitor cocktail 2 (P5726). Cell lysate was collected in a fresh Eppendorf tube. A total of 43.33 μl of 4× Laemmli sample buffer (Bio-Rad) (containing 10% β-mercaptoethanol) was added to the cell lysate and then heated at 95 °C for 10 min. The boiled samples were either used for western blot analysis or stored at –80 °C.

Western blotting

A total of 30 µl of protein lysate was loaded per well and separated on a 4–20% SDS-PAGE gel for 30 min at 250 V. Proteins were then transferred onto a nitrocellulose membrane for 12 min using a semi-dry turbo transfer apparatus (Bio-Rad). The membranes were blocked in 5% BSA for 1 h, after which they were probed overnight with respective primary antibodies on a rocker at 4 °C. The primary antibodies used in this assay were β-actin (1:10,000; Cell Signaling Technologies), p-ERK1/2 p44/42 (1:10,000; Cell Signaling Technologies) and p-AKT S473 (1:2,000; Cell Signaling Technologies). The next day, membranes were washed three times with 1× TBS-T and then incubated with anti-rabbit HRP conjugated secondary antibody (1:10,000; Bio-Rad) for 1 h. For p-AKT S473, following washes, the membrane was blocked in 5% milk at room temperature for 1 h and then incubated in the respective HRP-conjugated secondary antibody (1:2,000) prepared in 5% milk, for 1 h. They were developed using Immobilon Western chemiluminescent substrate (EMD Millipore), followed by quantification using NIH ImageJ analysis software. The raw scans of the western blot results are shown in Supplementary Fig. 5. Quantifications were done by calculating the peak area for each band. Inhibition curve fit and corresponding IC₅₀ values were determined using GraphPad Prism 9 software.

IL-7R α minibinder antagonist assay

HEK293T cells were cultured in DMEM medium with 10% FBS at 37 °C and 5% CO₂. Cells were co-transfected with 1,000 ng pcDNA3-γ common, 300 ng pMET7-HA-IL-7R α , 200 ng pMX-IRES-GFP-hJak3, 300 ng empty pMET7 vector and 200 ng pGL3-b-casein-luci STAT5 reporter plasmid per well of a 6-well plate. One day after transfection, cells were detached with cell dissociation buffer (Life Technologies), re-suspended in DMEM + 10% FCS and 2% of cells were seeded in 96-well plate as previously described⁷⁷ and stimulated overnight with 50 pM human IL-7 (Immunotools) and increasing concentrations of IL-7R α minibinder. STAT5-dependent luciferase activity was measured on the next day using a GloMax 96 microplate luminometer. The fold-induction of luciferase activity was calculated by the ratio of the luminescence signal from cells stimulated with IL-7 to the signal from the unstimulated cells. The data were plotted and fitted to a log inhibitor versus response curve in GraphPad Prism. The

pcDNA3-gamma common was a gift from J. C. Renauld (Faculty of Medicine and Dentistry, UC Louvain, Belgium) and the pMX-IRES-GFP-hJak3 vector⁸⁶ was provided by S. N. Constantinescu (Ludwig Institute for Cancer Research, Belgium). The pMET7-HA-IL-7R α , empty pMET7 and pGL3- β -casein-luci vectors were provided by F. Peelman (UGent, Belgium).

Apparent SC₅₀ estimation from FACS and next-generation sequencing

The Pear program⁸⁷ was used to assemble the fastq files from the next-generation sequencing (NGS) runs. Translated, assembled reads were matched against the ordered designs to determine the number of counts for each design in each pool.

The critical assumption to the fitting here is to assume that the yeast cells displaying a particular design will follow a modified version of the standard K_d equation relating fraction bound to concentration:

where $\text{fraction_collected}_i$ is the fraction of the yeast cells displaying design i that were collected, concentration is the target concentration for sorting, and $\text{SC}_{50,i}$ is the apparent SC_{50} of the design (the concentration where 50% of the cells would be collected).

The next assumption is that all designs have the same expression level on the yeast surface and that 100% of yeast cells express sufficiently well to be collected in the ‘expression’ gate (that is, the right population in Supplementary Fig. 7).

These two assumptions, although probably false, enable fitting of the data with only one free parameter per design and no global free parameters. The correct version of equation (1) for this experiment probably has a different shape and slope from a perfect sigmoid; the net effect of correcting this would be that all SC_{50} values are scaled by a constant factor (which would not affect the relative comparisons made here). It can be shown by analysing the data that different designs result in different expression levels on yeast (one can examine the fraction collected, for strong binders at concentrations for which binding should be saturated). The net result is that experimentally, equation (1) is multiplied by a constant between 0 and 1 for each design. This constant seems to range from 0.2 to 0.7. As such, when fitting the data, fraction collected, values above 0.2 are considered saturating. However, because the 0.2 mark may represent 90% collection for poorly expressing designs and 30% collection for strongly expressing designs, the resulting SC_{50} fits may vary by up to fivefold. The alternative is to try to estimate an expression level; however, this becomes increasingly difficult with weaker binders that never saturate the experiment.

Apparent SC₅₀ estimation from FACS and NGS: point estimates

The following equation may be used to determine the fraction collected, i , for a single design in a single sort:

\$\$\begin{array}{c} \rm{Fraction} \\ \rm{child} \\ \rm{proportion} \\ \rm{pool} \\ \rm{parent} \\ \rm{FACS} \\ \rm{collection} \\ \rm{fraction} \end{array} \end{array} \\ (2)

where $\text{fraction_collected}_i$ is the proportion of cells carrying design i that were collected during the sort, $\text{proportion_child_pool}_i$ is the proportion of the total NGS counts for design i from the pool that was collected, $\text{proportion_parent_pool}_i$ is the proportion of the total NGS counts for design i from the pool that was the input for the sorter, and FACS collection

fraction was the fraction of the yeast cells collected during the specific sort (a number extracted from the FACS machine itself).

This point-estimate method is best suited for asking which designs have $SC_{50} < SC_{50,0}$ by determining the expected $fraction_collected_i$ for a given sorting concentration and $SC_{50,0}$. The sorting concentration and $SC_{50,0}$ should be selected such that equation (1) results in an expected $fraction_collected_i$ less than 0.2 to circumvent the expression issues mentioned above. Then, any designs with $fraction_collected_i$ greater than the cut-off may say that their SC_{50} is less than $SC_{50,0}$. Designs with low numbers of counts are suspect, see the ‘Doubly transformed yeast cells’ section below. For this analysis, any designs with fewer than max possible passenger cells were eliminated.

This method may be applied to avidity sorts; however, the resulting SC_{50} would be the SC_{50} during avidity experiments. It is unclear what the precise mathematical effect of avidity is, and as such we do not compare avidity SC_{50} values with non-avidity SC_{50} values.

Apparent SC₅₀ estimation from FACS and NGS: doubly transformed yeast cells

Doubly transformed yeast cells represent a major source of error in these experiments. Although rare, a yeast cell that contains two plasmids, one of a strong binder and one of a non-binder, will carry the non-binder plasmid through the sorting process. The net result is that the non-binder will end up with counts that track the strong binder; however, at a greatly reduced absolute number. Note that rare is a relative term here. Although the odds of any two specific plasmids being in one cell is low, in the entire pool of yeast, doubly transformed cells seem to be common.

We chose to address this issue by making the following assumption: non-binders that take advantage of a doubly transformed yeast cell do so from precisely one double-transformation event. In other words, we assumed that the same non-binding plasmid did not get doubly transformed into two separate strong-binding yeast. This assumption allows us to estimate the

largest number of cells we would expect to see from a doubly transformed plasmid:

$$\begin{array}{c}
 \text{\rm M} \text{\rm a} \text{\rm x} \text{\rm _} \text{\rm p} \\
 \text{\rm o} \text{\rm s} \text{\rm s} \text{\rm i} \text{\rm b} \text{\rm l} \text{\rm e} \text{\rm _} \\
 \text{\rm p} \text{\rm a} \text{\rm s} \text{\rm s} \text{\rm e} \text{\rm n} \text{\rm g} \text{\rm e} \\
 \text{\rm r} \text{\rm _} \text{\rm c} \text{\rm e} \text{\rm l} \text{\rm l} \\
 \text{\rm s} = \frac{\text{\rm c} \text{\rm e} \text{\rm l} \text{\rm s} \text{\rm _}}{\text{\rm c} \text{\rm o} \text{\rm l} \text{\rm l} \text{\rm e} \text{\rm c} \text{\rm t} \text{\rm e} \\
 \text{\rm d} \text{\rm _} \text{\rm i} \text{\rm _} \text{\rm max} \text{\rm c} \text{\rm e} \text{\rm l} \text{\rm l} \text{\rm l} \\
 \text{\rm s} \text{\rm _} \text{\rm s} \text{\rm o} \text{\rm r} \text{\rm t} \text{\rm e} \text{\rm d} \\
 \text{\rm _} \text{\rm R} \text{\rm 1} \text{\rm _} \text{\rm i} \text{\rm _} \text{\rm max} \text{\rm \times} \text{\rm c} \text{\rm e} \\
 \text{\rm l} \text{\rm l} \text{\rm c} \text{\rm o} \text{\rm p} \text{\rm i} \text{\rm e} \\
 \text{\rm s} \text{\rm _} \text{\rm b} \text{\rm e} \text{\rm f} \text{\rm o} \text{\rm r} \text{\rm e} \\
 \text{\rm _} \text{\rm f} \text{\rm i} \text{\rm r} \text{\rm s} \text{\rm t} \text{\rm _} \text{\rm s} \\
 \text{\rm o} \text{\rm r} \text{\rm r} \text{\rm t} \text{\rm \end{array}}$$

(3)

where $\text{max_possible_passenger_cells}$ is the highest number of cells that we would expect a non-binding plasmid to occupy, $\text{cells_collected}_{i_max}$ is the number of cells collected in this round for the design with the greatest number of cells collected, $\text{cells_sorted_R1}_{i_max}$ is the number of cells sorted for design i_max (the same design from $\text{cells_collected}_{i_max}$), and cell copies before first sort is the number of copies of each cell that occurred before the first sort ($2^{\text{no. of cell divisions}}$). The number of cells_collected_i may be approximated by multiplying the number of cells the FACS machine collected by the proportion of the pool that design i represents. The number of cells_sorted_i may be estimated by either dividing the cells_collected_i by the FACS_collection_fraction or by multiplying the number of cells fed to the FACS machine by the proportion of design i in that pool.

With this number in hand, one can set a floor for the number of cells that one would expect to see. Any design with fewer than this number of cells cannot be considered for calculations because it is unclear whether or not that cell is part of a doubly transformed yeast cell. On the whole, this method reduces false-positive binders but also removes true-positive binders that did not

transform well. It is wise to simply drop designs from the downstream calculations that did not transform well.

Apparent SC₅₀ estimation from FACS and NGS: full estimate

Estimation of an upper and lower bound on the SC₅₀ from the data may be performed by looking at an arbitrary number of sorting experiments. Taking a $P(\text{SC}_{50} == \text{SC}_{50,0} | \text{data})$ and performing Bayesian analysis, one arrives at a confidence interval for the actual SC₅₀ value. This analysis may be performed at every sort and the resulting distributions combined to produce a robust estimate.

Each sort may be modelled as a binomial distribution where $P = \text{fraction_collected}$ from equation (1) using concentration = sorting_concentration and $\text{SC}_{50} = \text{SC}_{50,0}$; $n = \text{cells_sorted}_i$; and $x = \text{cells_collected}_i$. By performing this analysis at a range of SC_{50,0} values and examining the probability this could happen by the binomial distribution, one arrives at $P(\text{SC}_{50} == \text{SC}_{50,0} | \text{data})$. Specifically for this analysis, the cumulative distribution function (CDF) of the binomial was used with the null hypothesis that $\text{SC}_{50} == \text{SC}_{50,0}$.

Care should be taken for the valid range of P . As stated previously, it is wise to cap the expected value of P to 0.2 to account for expression levels and to floor the value such that $n \times P$ does not fall below max possible passenger cells. In our implementation, if x falls into a range that has been clipped, a probability of 1 is returned.

The code to perform this entire analysis is available in the [Supplementary Information](#).

SSM validation: relax protocol

To remove artefacts from designs and to discover the best orientation for each SSM mutation, all binders were relaxed using the Rosetta beta_nov16 score function before calculations began (30 replicates using 5 repeats of cartesian FastRelax taking the best scoring model). Relaxation of point

mutants then used the standard cartesian FastRelax procedure and allowed all residues within 10 Å of the mutation to relax. The backbone coordinates of those residues on the binder were allowed to relax while the target was held constant. The best of three (as evaluated by Rosetta energy) was chosen as the representative model. An xml is provided in the [Supplementary Information](#) to perform this relaxation.

SSM validation: entropy score

To validate that the designed binder was folded into the correct shape and was using its designed interface to bind to the target, the entropy of the interface, monomer core and monomer surface were examined. For each position on the binder, the sequence entropy (Shannon entropy) of each position was calculated using the observed frequencies of each amino acid in the NGS. The specific pool that was chosen for this analysis was the pool with concentration closest to tenfold lower than the calculated SC_{50} of the parent.

After the per-position sequence entropy was calculated, the average per-position entropy of the SASA-hidden positions contacting the target (interface core), the SASA-hidden positions not contacting the target (monomer core) and the fully exposed positions not contacting the target (monomer surface) were calculated. A simple subtraction was performed according to equation (4):

$$\begin{aligned} & \$\$ \{ \backslash rm\{ I \} \} \{ \backslash rm\{ n \} \} \{ \backslash rm\{ t \} \} \{ \backslash rm\{ e \} \} \{ \backslash rm\{ r \} \} \{ \backslash rm\{ m \} \} \{ \backslash rm\{ e \} \} \\ & \{ \backslash rm\{ d \} \} \{ \backslash rm\{ i \} \} \{ \backslash rm\{ a \} \} \{ \backslash rm\{ t \} \} \{ \backslash rm\{ e \} \} \backslash \{ \backslash rm\{ e \} \} \{ \backslash rm\{ n \} \} \{ \backslash rm\{ t \} \} \\ & \{ \backslash rm\{ r \} \} \{ \backslash rm\{ o \} \} \{ \backslash rm\{ p \} \} \{ \backslash rm\{ y \} \} \backslash \{ \backslash rm\{ s \} \} \{ \backslash rm\{ c \} \} \{ \backslash rm\{ o \} \} \{ \backslash rm\{ r \} \} \\ & \{ \backslash rm\{ e \} \} = \{ S \} _- \{ \backslash rm\{ m \} \} \{ \backslash rm\{ o \} \} \{ \backslash rm\{ n \} \} \{ \backslash rm\{ o \} \} \{ \backslash rm\{ m \} \} \{ \backslash rm\{ e \} \} \\ & \{ \backslash rm\{ r \} \} \{ \backslash rm\{ _ \} \} \{ \backslash rm\{ c \} \} \{ \backslash rm\{ o \} \} \{ \backslash rm\{ r \} \} \{ \backslash rm\{ e \} \} \} + \{ S \} _- \{ \backslash rm\{ i \} \} \\ & \{ \backslash rm\{ n \} \} \{ \backslash rm\{ t \} \} \{ \backslash rm\{ e \} \} \{ \backslash rm\{ r \} \} \{ \backslash rm\{ f \} \} \{ \backslash rm\{ a \} \} \{ \backslash rm\{ c \} \} \{ \backslash rm\{ e \} \} \\ & \{ \backslash rm\{ _ \} \} \{ \backslash rm\{ c \} \} \{ \backslash rm\{ o \} \} \{ \backslash rm\{ r \} \} \{ \backslash rm\{ e \} \} \} - \{ S \} _- \{ \backslash rm\{ m \} \} \{ \backslash rm\{ o \} \} \\ & \{ \backslash rm\{ n \} \} \{ \backslash rm\{ o \} \} \{ \backslash rm\{ m \} \} \{ \backslash rm\{ e \} \} \{ \backslash rm\{ r \} \} \{ \backslash rm\{ _ \} \} \{ \backslash rm\{ s \} \} \\ & \{ \backslash rm\{ u \} \} \{ \backslash rm\{ r \} \} \{ \backslash rm\{ f \} \} \{ \backslash rm\{ a \} \} \{ \backslash rm\{ c \} \} \{ \backslash rm\{ e \} \} \} \$\$ \end{aligned} \quad (4)$$

where S_{region} is the average entropy of that region.

Finally, the probability that the score could have come from totally random data was computed by performing the above calculation on the actual data, and then performing the same calculation 100 times, but randomly mismatching the observed counts among all SSM point mutations. In this way, the experimental noise is kept constant among the 100 decoy datasets.

The final step to arrive at a P value was to calculate the mean and standard deviation of the 100 decoy intermediate entropy scores and to find the P value with the Normal CDF function of the binder's intermediate entropy score.

SSM validation: Rosetta accuracy score

To further assess the accuracy of the design model, the correlation between the predicted effect on binding by Rosetta was compared with the experimental data. The effect from Rosetta can be broken into two components: monomer stabilization/destabilization and interface stabilization/destabilization. The effect on the monomer energy will affect the fraction of the proteins that are folded in solution. This fraction of folded proteins will then worsen the affinity because only the folded proteins are able to bind. The effect on the monomer stability was estimated by taking the difference in Rosetta energy between the native relaxed dock and the mutant relaxed dock and looking only at the change in Rosetta score of the docked protein (excluding energies arising from cross-interface edges). The effect on the target energy was calculated the same and was considered to directly affect the binding energy. The binding energy was calculated by taking the difference in Rosetta score between the docked and undocked conformations (but with no repacking or minimization in the unbound form). An xml exists in the [Supplementary Information](#) to perform this calculation.

The effect on the $P(\text{fold monomer})$ was estimated by first determining the predicted ΔG_{fold} of the native protein.

$$\begin{aligned}
 & \$\$P(\{\text{\rm fold}\}; \text{monomer}\}) = \{\exp\} (\frac{\Delta G_f - \Delta G_m}{kT}) \\
 & \Delta G_f = \Delta G_f^0 + \Delta G_{\text{bind}} \\
 & \Delta G_m = \Delta G_m^0 + \Delta G_{\text{bind}} \\
 & \Delta G_f^0 = \Delta G_f^0 + \Delta G_{\text{monomer}} \\
 & \Delta G_m^0 = \Delta G_m^0 + \Delta G_{\text{monomer}} \\
 & \Delta G_{\text{bind}} = \Delta G_{\text{bind}} + \Delta G_{\text{interface}}
 \end{aligned}$$

(5)

$$\begin{aligned}
 & \$\$ \Delta G_f^0 = kT \ln \left(\frac{P(\{\text{\rm fold}\}; \text{monomer}\})}{P(\{\text{\rm fold}\}; \text{native}\})} \right) \\
 & P(\{\text{\rm fold}\}; \text{monomer}\}) = \frac{e^{-\Delta G_f^0 / kT}}{1 + e^{-\Delta G_f^0 / kT}}
 \end{aligned}$$

(6)

Where k is the Boltzmann constant and T is temperature, which was set to 300 K for this calculation.

Using equations (5) and (6), the predicted ΔG_{fold} for the native design was estimated by performing a least-squares fit of all mutations that did not occur in residues at the interface. A rudimentary confidence interval was created by allowing all ΔG_{fold} values that resulted in a root mean squared error of within 0.25 kcal mol⁻¹ of the best ΔG_{fold} value. Typical confidence intervals spanned 3 kcal mol⁻¹.

$$\begin{aligned} \$\$ \Delta \{ \text{\rm dd} \} \{ G \}_{\text{R Rosetta}} = & \Delta \{ \text{\rm dd} \} \\ & \{ G \}_{\text{monomer effect}} + \Delta \{ \text{\rm dd} \} \{ G \}_{\text{i interface effect}} + \Delta \{ \text{\rm dd} \} \{ G \}_{\text{target effect}} \\ (7) \end{aligned}$$

With the ΔG_{fold} in hand, the predicted effect on the binding energy could be computed according to equation (7). The values of ΔG_{fold} inside the confidence range for ΔG_{fold} that produced the largest and smallest $\Delta ddG_{\text{Rosetta}}$ were used to produce a confidence interval for $\Delta ddG_{\text{Rosetta}}$.

The per-position accuracy was assessed by determining whether the confidence interval for $\Delta ddG_{\text{Rosetta}}$ was compatible with the confidence interval for the SC₅₀ from the experimental data. A buffer of 1 kcal mol⁻¹ was allowed.

With the per-position accuracies in hand, the overall percentage of mutations that Rosetta was able to explain in the monomer core and interface core was assessed. This produced an overall Rosetta accuracy score.

In the same way as the entropy score, 100 decoys with randomly shuffled SC₅₀ values were subjected to the same procedure. The mean and standard deviation of the decoys was determined and the P value for the Rosetta score was determined using the Normal CDF function.

Reporting summary

Further information on research design is available in the [Nature Research Reporting Summary](#) linked to this paper.

Data availability

The atomic coordinates and experimental data of H3_mb in complex with H3 HA, TrkA_mb in complex with TrkA, unbound FGFR2_mb, FGFR2_mb in complex with FGFR4, unbound IL-7Ra_mb, IL-7Ra_mb in complex with IL-7Ra and VirB8_mb in complex with VirB8 have been deposited in the RCSB PDB with the accession numbers [7RDH](#), [7N3T](#), [7N1K](#), [7N1J](#), [7S5B](#), [7OPB](#) and [7SH3](#), respectively. Diffraction images for the TrkA–minibinder complex have been deposited in the SBGrid Data Bank with the identifier [838](#). The Rosetta macromolecular modelling suite (<https://www.rosettacommons.org>) is freely available to academic and non-commercial users. Commercial licences for the suite are available through the University of Washington Technology Transfer Office.

Code availability

The Rosetta macromolecular modelling suite (<https://www.rosettacommons.org>) is freely available to academic and non-commercial users. Commercial licences for the suite are available through the University of Washington Technology Transfer Office. The design scripts and main PDB models, computational protocol for data analysis, experimental data and analysis scripts, the entire miniprotein scaffold library, all the design models and NGS results used in this paper can be downloaded from file servers hosted by the Institute for Protein Design: http://files.ipd.uw.edu/pub/robust_de_novo_design_minibinders_2021/supplemental_files/scripts_and_main_pdbs.tar.gz, http://files.ipd.uw.edu/pub/robust_de_novo_design_minibinders_2021/supplemental_files/computational_protocol_analysis.tar.gz, http://files.ipd.uw.edu/pub/robust_de_novo_design_minibinders_2021/supplemental_files/experimental_data_and_analysis.tar.gz, http://files.ipd.uw.edu/pub/robust_de_novo_design_minibinders_2021/supplemental_files/scaffolds.tar.gz, http://files.ipd.uw.edu/pub/robust_de_novo_design_minibinders_2021/supplemental_files/structures.tar.gz.

[lemental_files/design_models_pdb.tar.gz](http://files.ipd.uw.edu/pub/robust_de_novo_design_minibinders_2021/supplemental_files/design_models_pdb.tar.gz) and http://files.ipd.uw.edu/pub/robust_de_novo_design_minibinders_2021/supplemental_files/design_models_silent.tar.gz. All the files are stored in compressed gzip format. Once the files have been downloaded and decompressed, there is a detailed description of the binder design pipeline and the whole process can be reproduced based on those files. The source code for RIF docking implementation is freely available at <https://github.com/rifdock/rifdock>.

References

1. Chevalier, A. et al. Massively parallel de novo protein design for targeted therapeutics. *Nature* **550**, 74–79 (2017).
2. Strauch, E. M. et al. Computational design of trimeric influenza-neutralizing proteins targeting the hemagglutinin receptor binding site. *Nat. Biotechnol.* **35**, 667–671 (2017).
3. Silva, D. A. et al. De novo design of potent and selective mimics of IL-2 and IL-15. *Nature* **565**, 186–191 (2019).
4. Baran, D. et al. Principles for computational design of binding antibodies. *Proc. Natl Acad. Sci. USA* **114**, 10900–10905 (2017).
5. Fleishman, S. J. et al. Computational design of proteins targeting the conserved stem region of influenza hemagglutinin. *Science* **332**, 816–821 (2011).
6. Dou, J. et al. De novo design of a fluorescence-activating β-barrel. *Nature* **561**, 485–491 (2018).
7. Koga, N. et al. Principles for designing ideal protein structures. *Nature* **491**, 222–227 (2012).
8. Linsky, T. et al. Sampling of structure and sequence space of small protein folds. Preprint at *bioRxiv* <https://doi.org/10.1101/2021.03.10.434454> (2021).

9. Maguire, J. B. et al. Perturbing the energy landscape for improved packing during computational protein design. *Proteins* **89**, 436–449 (2021).
10. Rocklin, G. J. et al. Global analysis of protein folding using massively parallel design, synthesis, and testing. *Science* **357**, 168–175 (2017).
11. 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).
12. Brunette, T. J. et al. Modular repeat protein sculpting using rigid helical junctions. *Proc. Natl Acad. Sci. USA* **117**, 8870–8875 (2020).
13. Coventry, B. & Baker, D. Protein sequence optimization with a pairwise decomposable penalty for buried unsatisfied hydrogen bonds. *PLoS Comp. Biol.* **17**, e1008061 (2021).
14. Mackenzie, C. O., Zhou, J. & Grigoryan, G. Tertiary alphabet for the observable protein structural universe. *Proc. Natl Acad. Sci. USA* **113**, E7438–E7447 (2016).
15. Wiesmann, C., Ultsch, M. H., Bass, S. H. & de Vos, A. M. Crystal structure of nerve growth factor in complex with the ligand-binding domain of the TrkA receptor. *Nature* **401**, 184–188 (1999).
16. Plotnikov, A. N., Hubbard, S. R., Schlessinger, J. & Mohammadi, M. Crystal structures of two FGF–FGFR complexes reveal the determinants of ligand-receptor specificity. *Cell* **101**, 413–424 (2000).
17. Garrett, T. P. et al. Crystal structure of a truncated epidermal growth factor receptor extracellular domain bound to transforming growth factor α . *Cell* **110**, 763–773 (2002).
18. Shim, A. H. et al. Structures of a platelet-derived growth factor/propeptide complex and a platelet-derived growth factor/receptor complex. *Proc. Natl Acad. Sci. USA* **107**, 11307–11312 (2010).

19. Croll, T. I. et al. Higher-resolution structure of the human insulin receptor ectodomain: multi-modal inclusion of the insert domain. *Structure* **24**, 469–476 (2016).
20. Xu, Y. et al. How ligand binds to the type 1 insulin-like growth factor receptor. *Nat. Commun.* **9**, 821 (2018).
21. Barton, W. A. et al. Crystal structures of the Tie2 receptor ectodomain and the angiopoietin-2–Tie2 complex. *Nat. Struct. Mol. Biol.* **13**, 524–532 (2006).
22. McElroy, C. A., Dohm, J. A. & Walsh, S. T. Structural and biophysical studies of the human IL-7/IL-7R α complex. *Structure* **17**, 54–65 (2009).
23. Arnett, K. L., Harrison, S. C. & Wiley, D. C. Crystal structure of a human CD3- ϵ / δ dimer in complex with a UCHT1 single-chain antibody fragment. *Proc. Natl Acad. Sci. USA.* **101**, 16268–16273 (2004).
24. Radaev, S. et al. Ternary complex of transforming growth factor- β 1 reveals isoform-specific ligand recognition and receptor recruitment in the superfamily. *J. Biol. Chem.* **285**, 14806–14814 (2010).
25. Ekiert, D. C. et al. Cross-neutralization of influenza A viruses mediated by a single antibody loop. *Nature* **489**, 526–532 (2012).
26. Gillespie, J. J. et al. Structural insight into how bacteria prevent interference between multiple divergent type IV secretion systems. *mBio* **6**, e01867-15 (2015).
27. Birtalan, S. et al. The intrinsic contributions of tyrosine, serine, glycine and arginine to the affinity and specificity of antibodies. *J. Mol. Biol.* **377**, 1518–1528 (2008).
28. Lemmon, M. A. & Schlessinger, J. Cell signaling by receptor tyrosine kinases. *Cell* **141**, 1117–1134 (2010).

29. Markovic, I. & Savvides, S. N. Modulation of signaling mediated by TSLP and IL-7 in inflammation, autoimmune diseases, and cancer. *Front. Immunol.* **11**, 1557 (2020).
30. Webster, R. G., Bean, W. J., Gorman, O. T., Chambers, T. M. & Kawaoka, Y. Evolution and ecology of influenza A viruses. *Microbiol. Rev.* **56**, 152–179 (1992).
31. Nobusawa, E. et al. Comparison of complete amino acid sequences and receptor-binding properties among 13 serotypes of hemagglutinins of influenza A viruses. *Virology* **182**, 475–485 (1991).
32. Bullough, P. A., Hughson, F. M., Skehel, J. J. & Wiley, D. C. Structure of influenza haemagglutinin at the pH of membrane fusion. *Nature* **371**, 37–43 (1994).
33. Ekiert, D. C. et al. Antibody recognition of a highly conserved influenza virus epitope. *Science* **324**, 246–251 (2009).
34. Corti, D. et al. A neutralizing antibody selected from plasma cells that binds to group 1 and group 2 influenza A hemagglutinins. *Science* **333**, 850–856 (2011).
35. Joyce, M. G. et al. Vaccine-induced antibodies that neutralize group 1 and group 2 influenza A viruses. *Cell* **166**, 609–623 (2016).
36. Kadam, R. U. et al. Potent peptidic fusion inhibitors of influenza virus. *Science* **358**, 496–502 (2017).
37. van Dongen, M. J. P. et al. A small-molecule fusion inhibitor of influenza virus is orally active in mice. *Science* **363**, eaar6221 (2019).
38. Cao, L. et al. De novo design of picomolar SARS-CoV-2 miniprotein inhibitors. *Science* **370**, 426–431 (2020).
39. Case, J. B. et al. Ultrapotent miniproteins targeting the SARS-CoV-2 receptor-binding domain protect against infection and disease. *Cell Host Microbe* **29**, 1151–1161 (2021).

40. Quijano-Rubio, A. et al. De novo design of modular and tunable protein biosensors. *Nature* **591**, 482–487 (2021).
41. Frejd, F. Y. & Kim, K. T. Affibody molecules as engineered protein drugs. *Exp. Mol. Med.* **49**, e306 (2017).
42. Lim, Y. et al. GC1118, an anti-EGFR antibody with a distinct binding epitope and superior inhibitory activity against high-affinity EGFR ligands. *Mol. Cancer Ther.* **15**, 251–263 (2016).
43. Zhang, Y. & Skolnick, J. TM-align: a protein structure alignment algorithm based on the TM-score. *Nucleic Acids Res.* **33**, 2302–2309 (2005).
44. Silva, D. A., Correia, B. E. & Procko, E. Motif-driven design of protein–protein interfaces. *Methods Mol. Biol.* **1414**, 285–304 (2016).
45. Leaver-Fay, A. et al. ROSETTA3: an object-oriented software suite for the simulation and design of macromolecules. *Methods Enzymol.* **487**, 545–574 (2011).
46. Hoover, D. M. & Lubkowski, J. DNAWorks: an automated method for designing oligonucleotides for PCR-based gene synthesis. *Nucleic Acids Res.* **30**, e43 (2002).
47. Benatuil, L., Perez, J. M., Belk, J. & Hsieh, C. M. An improved yeast transformation method for the generation of very large human antibody libraries. *Protein Eng. Des. Sel.* **23**, 155–159 (2010).
48. Stevens, J. et al. Structure of the uncleaved human H1 hemagglutinin from the extinct 1918 influenza virus. *Science* **303**, 1866–1870 (2004).
49. Divine, R. et al. Designed proteins assemble antibodies into modular nanocages. *Science* **372**, eabd9994 (2021).
50. Xu, Y. et al. Addressing polyspecificity of antibodies selected from an in vitro yeast presentation system: a FACS-based, high-throughput selection and analytical tool. *Prot. Eng. Des. Sel.* **26**, 663–670 (2013).

51. Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326 (1997).
52. McCoy, A. J. et al. Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674 (2007).
53. McCoy, A. J., Grosse-Kunstleve, R. W., Storoni, L. C. & Read, R. J. Likelihood-enhanced fast translation functions. *Acta Crystallogr. D* **61**, 458–464 (2005).
54. Adams, P. D. et al. PHENIX: building new software for automated crystallographic structure determination. *Acta Crystallogr. D* **58**, 1948–1954 (2002).
55. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D* **60**, 2126–2132 (2004).
56. Chen, V. B. et al. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D* **66**, 12–21 (2010).
57. Wehrman, T. et al. Structural and mechanistic insights into nerve growth factor interactions with the TrkA and p75 receptors. *Neuron* **53**, 25–38 (2007).
58. Kabsch, W. XDS. *Acta Crystallogr. D* **66**, 125–132 (2010).
59. Legrand, P. XDSME: XDS Made Easier. *GitHub* <https://github.com/legrandp/xdsme> (2017)
60. Evans, P. R. An introduction to data reduction: space-group determination, scaling and intensity statistics. *Acta Crystallogr. D* **67**, 282–292 (2011).
61. Evans, P. R. & Murshudov, G. N. How good are my data and what is the resolution? *Acta Crystallogr. D* **69**, 1204–1214 (2013).
62. Winn, M. D. et al. Overview of the CCP4 suite and current developments. *Acta Crystallogr. D* **67**, 235–242 (2011).

63. Terwilliger, T. C. et al. Iterative model building, structure refinement and density modification with the PHENIX AutoBuild wizard. *Acta Crystallogr. D* **64**, 61–69 (2008).
64. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D* **66**, 486–501 (2010).
65. Adams, P. D. et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D* **66**, 213–221 (2010).
66. Echols, N. et al. Graphical tools for macromolecular crystallography in PHENIX. *J. Appl. Crystallogr.* **45**, 581–586 (2012).
67. Afonine, P. V. et al. Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallogr. D* **68**, 352–367 (2012).
68. Painter, J. & Merritt, E. A. Optimal description of a protein structure in terms of multiple groups undergoing TLS motion. *Acta Crystallogr. D* **62**, 439–450 (2006).
69. Headd, J. J. et al. Flexible torsion-angle noncrystallographic symmetry restraints for improved macromolecular structure refinement. *Acta Crystallogr. D* **70**, 1346–1356 (2014).
70. Karplus, P. A. & Diederichs, K. Linking crystallographic model and data quality. *Science* **336**, 1030–1033 (2012).
71. Morin, A. et al. Collaboration gets the most out of software. *eLife* **2**, e01456 (2013).
72. Plotnikov, A. N., Schlessinger, J., Hubbard, S. R. & Mohammadi, M. Structural basis for FGF receptor dimerization and activation. *Cell* **98**, 641–650 (1999).
73. Schlessinger, J. et al. Crystal structure of a ternary FGF–FGFR–heparin complex reveals a dual role for heparin in FGFR binding and

- dimerization. *Mol. Cell* **6**, 743–750 (2000).
74. Kabsch, W. Integration, scaling, space-group assignment and post-refinement. *Acta Crystallogr. D* **66**, 133–144 (2010).
75. Liebschner, D. et al. Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix. *Acta Crystallogr. D* **75**, 861–877 (2019).
76. Williams, C. J. et al. MolProbity: more and better reference data for improved all-atom structure validation. *Protein Sci.* **27**, 293–315 (2018).
77. Verstraete, K. et al. Structure and antagonism of the receptor complex mediated by human TSLP in allergy and asthma. *Nat. Commun.* **8**, 14937 (2017).
78. BUSTER v.2.10.2 (Global Phasing Ltd., 2016).
79. Joosten, R. P., Long, F., Murshudov, G. N. & Perrakis, A. The PDB_REDQ server for macromolecular structure model optimization. *IUCrJ* **1**, 213–220 (2014).
80. Evans, P. Scaling and assessment of data quality. *Acta Crystallogr. D* **62**, 72–82 (2006).
81. Jumper, J. et al. Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).
82. Berman, H. M. et al. The Protein Data Bank. *Nucleic Acids Res.* **28**, 235–242 (2000).
83. Berman, H., Henrick, K. & Nakamura, H. Announcing the worldwide Protein Data Bank. *Nat. Struct. Biol.* **10**, 980 (2003).
84. Grabowski, M. et al. A public database of macromolecular diffraction experiments. *Acta Crystallogr. D* **72**, 1181–1193 (2016).

85. Grabowski, M. et al. The Integrated Resource for Reproducibility in Macromolecular Crystallography: experiences of the first four years. *Struct. Dyn.* **6**, 064301 (2019).
86. Hornakova, T. et al. Acute lymphoblastic leukemia-associated JAK1 mutants activate the Janus kinase/STAT pathway via interleukin-9 receptor α homodimers. *J. Biol. Chem.* **284**, 6773–6781 (2009).
87. Zhang, J., Kober, K., Flouri, T. & Stamatakis, A. PEAR: a fast and accurate Illumina Paired-End reAd mergeR. *Bioinformatics* **30**, 614–620 (2014).
88. Buchan, D. W. A. & Jones, D. T. The PSIPRED Protein Analysis Workbench: 20 years on. *Nucleic Acids Res.* **47**, W402–W407 (2019).
89. Lauer, T. M. et al. Developability index: a rapid in silico tool for the screening of antibody aggregation propensity. *J. Pharm. Sci.* **101**, 102–115 (2012).

Acknowledgements

This work was supported by DARPA Synergistic Discovery and Design (SD2) HR0011835403 contract FA8750-17-C-0219 (to L.C., B.C., S.H. and D.B.); The Audacious Project at the Institute for Protein Design (to L.K.); the Open Philanthropy Project Improving Protein Design Fund (to B.C. and D.B.); funding from Eric and Wendy Schmidt by recommendation of the Schmidt Futures programme (to I.G. and L.M.); an Azure computing resource gift for COVID-19 research provided by Microsoft (to L.C. and B.C.); the National Institute of Allergy and Infectious Diseases (HHSN272201700059C to D.B., B.H. and L.S.; NIH R01 AI140245 to E.-M.S.; NIH R01 AI150855 to I.A.W.); the National Institute on Aging (R01AG063845 to B.H. and D.B.); the Defense Threat Reduction Agency (HDTRA1-16-C-0029 to D.B. and E.-M.S.); The Donald and Jo Anne Petersen Endowment for Accelerating Advancements in Alzheimer’s Disease Research (to N.B.); a gift from Gates Ventures (to M.D.); The Human Frontier Science Program (to A.Y.); and The Howard Hughes Medical Research Institute (K.M.J., K.C.G. and D.B.). Use of SSRL at

Stanford Linear Accelerator Center (SLAC) National Accelerator Laboratory is supported by the US Department of Energy (DOE) Office of Science, Office of Basic Energy Sciences under contract DE-AC02-76SF00515. The SSRL Structural Molecular Biology Program is supported by the DOE, Office of Biological and Environmental Research and the National Institutes of Health, National Institute of General Medical Sciences (including P41GM103393). A part of this work is based on research conducted at the Northeastern Collaborative Access Team beamlines, which are funded by the National Institute of General Medical Sciences from the National Institutes of Health (P30 GM124165). The Eiger 16M detector on the 24-ID-E beam line is funded by a NIH-ORIP HEI grant (S10OD021527). S.T.R.W. was supported by the CCR intramural research programme of NCI-NIH. GM/CA at the Advanced Photon Source at Argonne National Laboratory has been funded by the National Cancer Institute (ACB-12002) and the National Institute of General Medical Sciences (AGM-12006, P30GM138396). This research used resources of the Advanced Photon Source, a US DOE Office of Science User Facility operated for the DOE Office of Science by Argonne National Laboratory under contract no. DE-AC02-06CH11357. The Eiger 16M detector at GM/CA-XSD was funded by NIH grant S10 OD012289. S.N.S. acknowledges research support from Research Foundation Flanders (grants G0C2214N and G0E1516N) and the Hercules Foundation (no. AUGE-11-029). S.N.S. is a principal investigator of the VIB (Belgium). SSGCID is funded by federal funds from the National Institute of Allergy and Infectious Diseases (NIAID), National Institutes of Health (NIH), Department of Health and Human Services, under contract no. HHSN272201700059C from 1 September 2017. APS/LSCAT research used resources of the Advanced Photon Source, a US DOE Office of Science user facility operated for the DOE Office of Science by Argonne National Laboratory under contract no. DE-AC02- 06CH11357. Use of LS-CAT Sector 21 was supported by the Michigan Economic Development Corporation and the Michigan Technology Tri-Corridor (grant 085P1000817). We thank the staff of beamline ID23-2 (ESRF) for technical support and beamtime allocation; G. Ueda for providing the ANG1 protein for the TrkA competition assay; D. H. Fuller for providing the FI6v3 antibody for the HA competition assay; Y.-J. Park, A. Walls and D. Veesler for their collaborative research and cryo-EM structure determination for

minibinders targeting the SARS-CoV-2 spike; and K. Van Wormer and A. Curtis Smith for their laboratory support during COVID-19.

Author information

Author notes

1. These authors contributed equally: Longxing Cao, Brian Coventry

Affiliations

1. Department of Biochemistry, University of Washington, Seattle, WA, USA

Longxing Cao, Brian Coventry, Inna Goreshnik, Buwei Huang, William Sheffler, Nathaniel Bennett, Ashish Phal, Lisa Kozodoy, Michelle DeWitt, Lauren Miller, Asim K. Bera, Wei Yang, Lance Stewart, Hannele Ruohola-Baker & David Baker

2. Institute for Protein Design, University of Washington, Seattle, WA, USA

Longxing Cao, Brian Coventry, Inna Goreshnik, Buwei Huang, William Sheffler, Nathaniel Bennett, Lisa Kozodoy, Michelle DeWitt, Lauren Miller, Asim K. Bera, Wei Yang, Lance Stewart & David Baker

3. Molecular Engineering Graduate Program, University of Washington, Seattle, WA, USA

Brian Coventry & Nathaniel Bennett

4. Department of Bioengineering, University of Washington, Seattle, WA, USA

Buwei Huang & Ashish Phal

5. Department of Pharmacology, Yale University School of Medicine,
New Haven, CT, USA

Joon Sung Park, Joseph Schlessinger & Sangwon Lee

6. Howard Hughes Medical Institute, Stanford University School of
Medicine, Stanford, CA, USA

Kevin M. Jude, Aerin Yang, Lora Picton & K. Christopher Garcia

7. Department of Structural Biology, Stanford University School of
Medicine, Stanford, CA, USA

Kevin M. Jude, Aerin Yang, Lora Picton & K. Christopher Garcia

8. Department of Molecular and Cellular Physiology, Stanford University
School of Medicine, Stanford, CA, USA

Kevin M. Jude, Aerin Yang, Lora Picton & K. Christopher Garcia

9. VIB-UGent Center for Inflammation Research, Ghent, Belgium

Iva Marković, Koen H. G. Verschueren, Kenneth Verstraete & Savvas
N. Savvides

10. Unit for Structural Biology, Department of Biochemistry and
Microbiology, Ghent University, Ghent, Belgium

Iva Marković, Koen H. G. Verschueren, Kenneth Verstraete & Savvas
N. Savvides

11. Department of Integrative Structural and Computational Biology, The
Scripps Research Institute, La Jolla, CA, USA

Rameshwar U. Kadam, Steffen Bernard & Ian A. Wilson

12. Chemical Biology Laboratory, National Cancer Institute, National
Institutes of Health, Frederick, MD, USA

Scott Thomas Russell Walsh

13. J.A.M.E.S. Farm, Clarksville, MD, USA

Scott Thomas Russell Walsh

14. Institute for Stem Cell and Regenerative Medicine, University of Washington, Seattle, WA, USA

Ashish Phal & Hannele Ruohola-Baker

15. Department of Pharmaceutical and Biomedical Sciences, University of Georgia, Athens, GA, USA

Eva-Maria Strauch

16. UCB Pharma, Bainbridge Island, WA, USA

Nicholas D. DeBouver

17. Seattle Structural Genomics Center for Infectious Disease (SSGCID), Seattle, WA, USA

Nicholas D. DeBouver, Allison Pires & Bradley Hammerson

18. Seattle Children's Center for Global Infectious Disease Research, Seattle, WA, USA

Allison Pires

19. Department of Electrical and Computer Engineering, University of Washington, Seattle, WA, USA

Samer Halabiya

20. The Skaggs Institute for Chemical Biology, The Scripps Research Institute, La Jolla, CA, USA

Ian A. Wilson

21. Howard Hughes Medical Institute, University of Washington, Seattle, WA, USA

David Baker

Contributions

L.C., B.C. and D.B. designed the research. L.C. and B.C. contributed equally. L.C. and B.C. developed the binder design pipeline. W.S. developed the RIF docking method. L.C., B.C. and E.-M.S. designed the scaffold library. L.C., B.C., B. Huang and N.B. designed the binders. L.C., B.C., I.G., B. Huang, N.B., L.K., M.D., L.M., S.H. and W.Y. performed the yeast screening, expression and binding experiments. R.U.K., S.B. and I.A.W. prepared the H3 protein and solved the structure of the H3_mb complex. L.P., K.M.J. and A.Y. prepared the target protein, solved the structure of the complex and performed the competition assay for TrkA. J.S.P., J.S. and S.L. solved the structure of the FGFR2_mb complex. A. Phal performed the competition assay for FGFR2 and EGFR. I.M., K.H.G.V., K.V. and S.N.S. performed the IL-7R α competition assay and solved the structure of the IL-7R α _mb complex. S.T.R.W. solved the structure of the unbound IL-7R α _mb. B. Hammerson, N.D.D., A.P. and A.K.B. prepared the VirB8 target protein and solved the structure of the complex. All authors analysed data. L.S., I.A.W., H.R.-B., J.S., S.L., S.N.S., K.C.G. and D.B. supervised research. L.C., B.C. and D.B. wrote the manuscript with the input from the other authors. All authors revised the manuscript.

Corresponding author

Correspondence to [David Baker](#).

Ethics declarations

Competing interests

L.C., B.C., I.G., B.H., N.B., E.-M.S., L.S. and D.B. are co-inventors on a provisional patent application (21-0753-US-PRO) that incorporates discoveries described in this manuscript.

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 Detailed flow chart of the de novo miniprotein binder design pipeline.](#)

The computational design steps are colored as light green and experimental characterization and optimization steps are colored as light blue.

[Extended Data Fig. 2 Analysis of the critical steps of the de novo binder design pipeline.](#)

a, Comparison of the two docking approaches based on Rosetta ddG and contact molecular surface. Average and per-target distribution of the top 1% of binders in two key metrics after pooling equal-CPU-time dock-and-design trajectories. RifDock seeded with PatchDock outputs generated 300 outputs per scaffold that were trimmed to a total of 19,500 docks with “The Predictor” and designed using combinatorial side-chain optimization (orange). RifDock using the Hierarchical docking search generated 300 outputs per scaffold that were trimmed to a total of 19,500 docks with “The

Predictor” and subsequently designed (purple). Rosetta ddG refers to the predicted binding energy as calculated by Rosetta and Contact MS to key residues refers to the Contact Molecular Surface value (a distance weighted interfacial area calculation) to the key hydrophobic residues on the target that define this binding site. **b**, The rapid pre-screening method enriches docks with better Rosetta ddG and contact molecular surface. Average and per-target distribution of the top 1% of binders in two key metrics after pooling equal-CPU-time dock-and-design trajectories. The top 30 PatchDock outputs for the 1,000 helical scaffolds tested were designed using the RosettaScripts protocol (blue). The top 300 PatchDock outputs for the 1,000 helical scaffolds tested were trimmed to 21,000 with “The Predictor” and subsequently designed (red). **c**, The improved sequence design protocol yielded amino acid sequences more strongly predicted to fold to the monomer structure. The effect on fragment quality and Rosetta Score with different fragment-quality-guidance approaches. Rosetta using FastDesign with the standard LayerDesign settings was used to design 1,000 3-helical and 1,000 4-helical mini-protein scaffolds (blue). The same protocol was supplanted with the ConsensusLoopDesign TaskOperation (orange). The structure-based PSSM was used as an energy term in addition to the Standard Rosetta protocol (green). Two predictors of sequence-structure correspondence were found to improve without negatively affecting the computed Rosetta score of the binders. The probability that the designed sequence encoded for the wrong secondary structure was computed using PsiPred4⁸⁸ (left), and for each 9aa fragment of the designed scaffold, the closest match to a fragment in the Protein Data Bank with the same sequence was computed and averaged over the entire structure¹⁰ (center). Details can be found in the Supplemental Information. **d**, The improved sequence design protocol yielded amino acid sequences more strongly bound to the target. 10,000 scaffolds docked against the N-terminal domain of EGFR were designed with the RosettaScripts protocol while varying only the weight of the ProteinProteinInterfaceUpweighter. This TaskOperation multiplies all energies across the interface by the listed value during packing-design calculations.

Extended Data Fig. 3 Comparison of the native binding partners and the computational design models.

Side-by-side comparison of the native binding partners of the selected targets and the binding configurations of the computational designed models.

Extended Data Fig. 4 Biolayer interferometry characterization of binding of optimized designs to the corresponding targets.

Two-fold serial dilutions were tested for each binder and the highest concentration is labeled. For H3, TrkA, FGFR2, EGFR, PDGFR, IL-7Ra, CD3δ, TGF-β and VirB8, the biotinylated target proteins were loaded onto the Streptavidin (SA) biosensors, and incubated with miniprotein binders in solution to measure association and dissociation. For IGF1R and Tie2, MBP- (maltose binding protein) tagged miniprotein binders were used as the analytes. For InsulinR, the miniprotein binder was immobilized onto the Amine Reactive Second-Generation (AR2G) Biosensors and the insulin receptor was used as the analyte. The gray color represents experimental data and orange color represents fit curves. The fitting curves are poor at high binder concentrations due to the self- association of the binders through the interface hydrophobic residues, so we only kept the traces and fits at low binder concentrations.

Extended Data Fig. 5 Average SSM sequence entropy for different regions of binders.

The sequence entropy of a single position was calculated by looking at the counts from the sort with the concentration closest to 10-fold lower than the estimated parent SC_{50} and performing a simple Shannon entropy calculation on all amino acids observed at that position. Each plotted point is the average entropy of all positions within each of the three zones respectively. Validated vs Not Validated refers to the SSM Validation procedure with a cutoff of 0.005 (see [Methods](#) and [Extended Data Figure 15](#)). Since one would expect the core residues of the monomer and core residues of the interface to be conserved while the surface residues should not matter, the validated binders trend above the line. Points on the line do not show a difference between their surfaces and cores, potentially

indicating unfolded or misfolded proteins. Points below the line may be misfolded or binding with alternate residues.

Extended Data Fig. 6 Computational analysis of the experimental SSM results.

a, Ability of Rosetta to predict mutational effects. This graph shows the observed experimental effect of each mutation versus Rosetta's expected effect. For each plotted point, the delta refers to the effect versus the parent SSM design; therefore a "Beneficial" mutation is one that would improve affinity relative to the original designed protein the SSM was based on. The Δ Experimental ddg is derived from FACS data using the SC_{50} values (see [Methods](#)). Confidence intervals were collapsed to their center point to make this graph and "No effect" refers to mutations with less than a 1 kcal/mol change. Binder region definitions: Interface Core: residue contacts target protein and has no SASA (Solvent Accessible Surface Area) in bound state; Interface Boundary: residue contacts target protein, but does have SASA; Monomer Core: residue has no SASA and does not contact target; Monomer Boundary: residue has intermediate SASA and does not contact target; Monomer Surface: residue has full SASA and does not contact target. see [Methods](#) SSM Validation for further explanation. **b**, Mutations observed in SSM experiments that improved affinity bind at least 1kcal/mol graphed by relative frequency. Plotted is the $\# \text{ times Native_to_Mutant_improved_affinity} / \# \text{ times Native_to_Mutant_tested_in SSMs}$. A value of 0.10 with x-axis F and y-axis W could therefore represent that for 2 of 20 times W was substituted for Y, the affinity improved. Separated bars on each axis represent pooled data for the entire row/column. Grey boxes indicate mutations that occurred fewer than 5 times. Only SSM designs with a validation score of 0.005 or better were considered. While some cells are clipped, none extended beyond 0.25. Binder region definitions: Interface Core: residue contacts target protein and has no SASA in bound state; Interface Boundary: residue contacts target protein, but does have SASA; Monomer Core: residue has no SASA and does not contact target; Monomer Boundary: residue has intermediate SASA and does not contact target; Monomer Surface: residue has full SASA and does not contact target.

Extended Data Fig. 7 Competition experiments indicated the miniprotein binders bound to the targeted region.

Yeast cells displaying the TrkA binder (**a**), InsulinR binder (**b**), IGF1R binder (**c**), PDGFR binder (**d**) and Tie2 binder (**e**) were incubated with the target protein in the presence or absence of the native ligand as the competitor, and target protein binding to cells (y axis) was monitored with flow cytometry.

Extended Data Fig. 8 Inhibition of the TrkA miniprotein binder on the native TrkA-NGF signaling pathway.

a, Titration curves of nerve growth factor (NGF) on TrkA signaling in the presence of different concentrations of the TrkA miniprotein binder. The TrkA miniprotein binder shifted the IC₅₀ values of the TrkA response to NGF. **b**, The TrkA miniprotein binder showed no effects on the cell viability. TF-1 cells were treated with different concentrations of the TrkA miniprotein binder and the cell viability was quantified at both 24 and 48 hr. The mean values were calculated from duplicates for the pERK and pAKT signaling data, and triplicates for the cell proliferation and cell toxicity data. The error bars for the cell proliferation and cell cell toxicity data represent standard deviations.

Extended Data Fig. 9 Experimental characterization of the effects of the FGFR2 minibinder and the EGFR n-side minibinder on their native signaling.

a, FGFR2 mini binder (FGFR2_mb) inhibits FGF-induced ERK phosphorylation. Western Blot analysis showing reduction in FGF signaling (lanes 4-8) with increase in mini binder concentration. Lanes 3-4 show that EGF-induced ERK phosphorylation is unaffected by FGFR2 mini binder, eliminating any cross talk between the two receptors. **b**, EGFR n-side mini binder (EGFRn_mb) inhibits EGF-induced ERK and AKT phosphorylation. Western Blot analysis showing reduction in EGF signaling (lanes 4-8) with increase in mini binder concentration. Lanes 3-4 show that β FGF-induced ERK phosphorylation is unaffected by EGFR mini binder, eliminating any

crosstalk between the two receptors. **c**, Titration curve for bFGF mediated pERK signaling. (upper) Western Blot showing dose-dependent increase in FGF signaling with β FGF concentration. (lower) n = 2 biologically independent experimental repeats were performed, and quantification was done using ImageJ analysis software. The selected concentration for competition assays was 0.75 nM. **d**, Titration curve for EGF mediated pERK/pAkt signaling. (upper) Western Blot showing dose-dependent increase in EGF signaling with EGF concentration. (lower) n = 2 biologically independent experimental repeats were performed, and quantification was done using ImageJ analysis software. The selected concentration for competition assays was 1 nM. **e**, Representative Western Blot for inhibition curves – FGFR2 minibinder. Western Blot shows dose-dependent reduction in pERK signaling with mini minder concentration. Quantification was done using ImageJ analysis software. **f**, Representative Western Blot for inhibition curves – EGFR n-side minibinder. Western Blot shows dose-dependent reduction in (upper) pERK signaling and (lower) pAkt signaling with minibinder concentration. Quantification was done using ImageJ analysis software. **g**, Dose-dependent reduction in pAkt signaling elicited by 1 nM EGF in HUVECs with increase in EGFR n-side minibinder concentration. The IC₅₀ was calculated using a four-parameter-logistic equation in GraphPad Prism 9 software.

Extended Data Fig. 10 De novo design and experimental characterization of the influenza hemagglutinin (HA) binder.

a, Structure comparison of the stem region of group 1 HA and group 2 HA. The stem regions of H1 HA (A/Puerto Rico/8/1934) (left, PDB code: 1RU7) and H3 HA (A/Hong Kong/1/1968) (right, PDB code: 4WE4) are shown in cartoon and colored in pale cyan and pale green respectively, the key residues in the stem region are shown as sticks. Three major differences make the H3 HA stem region a more challenging target for designing de novo protein binders: the H3 HA stem region contains more polar residues and is more hydrophilic. Residues in H1 HA that are hydrophobic residues or small polar residues while the corresponding residues are polar or larger polar residues are highlighted in dashed circles; Trp21 adopts different configurations in H1 HA and H3 HA, and the targeting groove in H3 HA is much shallower and less hydrophobic; the H3 HA is glycosylated at Asn38,

and the carbohydrate side chains cover the hydrophobic groove and protect the HA stem region from binding by antibodies or designed binders. The insert shows a more extended view of the Asn38 glycosylation site on H3 HA. **b**, Binding of H3 binder to the H3 HA (A/Hong Kong/1/1968) N38D mutant (left) and H1 HA (A/Puerto Rico/8/1934) (right) with BLI. Two-fold serial dilutions were tested for each binder and the highest concentrations and the binder affinities are labeled. The gray color represents experimental data and orange color represents fit curves. **c**, The FI6v3 antibody competes with the binder for binding to the influenza A H1 hemagglutinin (left) and influenza A H3 hemagglutinin (right). Yeast cells displaying the H3 binder were incubated with 10 nM H1 or H3 in the presence or absence of 2 μ M FI6v3 antibody, and hemagglutinin binding to cells (y axis) was monitored with flow cytometry.

Extended Data Fig. 11 Structure characterization of the miniprotein binders without the target proteins.

Superimposition of the computation of the design model (silver) and the crystal structure for the FGFR2 binder (**a**) and IL-7R α (**b**) binder. The crystal structures of the miniprotein binders were determined without the target protein.

Extended Data Fig. 12 Analysis of the determinants of the success rate of de novo binder design.

a, Correlation between success rate and root mean square deviation (RMSD) with scaffolds. In this experiment, the accuracy of the scaffold library was examined with an experiment similar to Chevalier et al¹. The binding residues from known-good interfaces were copied onto scaffolds that closely resembled the known-good binders. If the scaffold folded properly and displayed these binding residues similarly to the original known-good interface, the hypothesis was that the scaffold would bind. This experiment sought to determine both the required accuracy of displayed sidechains to create a successful binder as well as to probe the accuracy of the scaffold library. If for instance, the scaffold library was perfectly accurate, this graph would indicate that if the C α RMSD of the

displayed sidechains deviates from the known-good conformation by 0.5 Å, that there would be a 15% chance of binding due to the intrinsic accuracy of sidechains required for binding. The scaffold library is likely not perfectly accurate however; as such, the correct interpretation would be: If the C α RMSD of the displayed sidechains according to the scaffold PDB model (which may not be perfectly correct) deviates by 0.5 Å C α RMSD, there is a 15% chance of binding. This 15% chance of binding arises in part from the likelihood that the scaffold will fold correctly and in part from the intrinsic required accuracy of sidechain placements for binding. Notably, the RMSD reported in this graph is far lower than the determined crystallographic accuracy of the IL-7R α binder when aligned by the receptor (the two interfacial helices are 1.5 Å C α RMSD when aligned by the IL-7R α receptor); however, if the two interfacial helices are aligned without regard for the receptor (the same calculation performed in this figure (i.e. the helices are superimposed on top of each other)) the C α RMSD is 0.43 Å. As such, the best explanation for this data is as follows: Although the predicted binding conformation of the complex structure was only accurate to 1.5 Å, the predicted monomer structure was correct to 0.43 Å. The comparison between scaffold and known-good interface was performed at the monomer level, and therefore, these new binders were successful because they assumed the correct monomer structure, which displayed the sidechains the same as the known-good binder, and therefore were able to bind, even though the known-good complex structure was not as accurate. This graph continues to show increased signal below 0.43 Å probably because the scaffolds at very low RMSD ended up being slightly structurally different for the same reason as the known-good binder. (i.e. if we crystallized one of the scaffolds that differed only by 0.2 Å, we would likely find that scaffold model and the scaffold crystal structure deviate by about 0.43 Å and that the scaffold crystal structure and the known-good crystal structure are very similar). **Method:** 11 IL-7R α SSM-validated interfaces were used as a starting point to create 2-helical grafts. All grafts consisted of 2-helices joined with a loop and the scaffold library was superimposed onto these two helices and the RMSD of the match was assessed. If a good match was found, the sidechains making strong interactions with IL-7R α were copied onto the scaffold and the remaining positions near the interface were allowed to redesign to avoid clashes. Plotted on the x-axis is the RMSD of the superposition of the 2-helices + loop between the motif and the scaffold.

The y-axis represents the fraction of binders with predicted $SC_{50s} < 3 \mu M$ with the number on top representing the denominator. **b**, Target success rate versus hydrophobicity. The y-axis shows what percentage of tested binders against the indicated target showed SC_{50} below 4 μM . The x-axis shows the hydrophobicity of the target region in SAP⁸⁹ units. A greater Δsap_score indicates greater hydrophobicity. While this graph is not completely fair as the authors improved the method with time, the trend is striking and can be used to estimate the difficulty of potential future targets. (The Δsap_score can be calculated on the target structure alone by observing the SAP score of all residues a potential binder would cover.).

Extended Data Fig. 13 Power of computational metrics to predict binders.

On the fully-relaxed binder dataset (see [Methods](#)), the ability of several computational metrics to predict which binders would have SC_{50} below 4 μM was assessed. In black and in the bar charts, data for all targets were pooled together. The bar charts show the success rate in each of the 10 percentiles for the metric while the black solid line shows the ROC plot for the metric. Each of the colored lines represents the correlation of this metric on each of the targets individually. The AUC of the overall black line is given in the upper left with the median of the AUC of the colored lines given immediately below.

Extended Data Table 1 Number of binders against the 12 targets as estimated from FACS sorting

Extended Data Table 2 Crystallographic data collection and refinement statistics

Supplementary information

Supplementary Information

Supplementary Figs. 1–8, Supplementary Table 1 and information for downloading the raw design models and design scripts.

Reporting Summary

41586 2022 4654 MOESM3 ESM.gz

Supplementary Code Design scripts and main pdb files. This compressed file is the main supplement and contains the following files: => cao_2021_protocol/ => design_models_final_combo_optimized/ => design_models_sequence/ => design_models_ssm_natives/ => ngs_analysis_scripts/

Supplementary Data 1 Experimental data and analysis. This file contains all the experimental results and the analysis protocols.

41586 2022 4654 MOESM5 ESM.gz

Supplementary Data 2 Computational protocol for data analysis. This file contains all the computational analysis we did for Fig. 1 and the Supplementary Figs. There are no experimental data here.

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

Cao, L., Coventry, B., Goreshnik, I. *et al.* Design of protein-binding proteins from the target structure alone. *Nature* **605**, 551–560 (2022). <https://doi.org/10.1038/s41586-022-04654-9>

- Received: 28 September 2021
- Accepted: 15 March 2022
- Published: 24 March 2022
- Issue Date: 19 May 2022
- DOI: <https://doi.org/10.1038/s41586-022-04654-9>

Share this article

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

[Get shareable link](#)

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

[Copy to clipboard](#)

Provided by the Springer Nature SharedIt content-sharing initiative

Further reading

- [**Targeted design of protein binders**](#)

- Sarah Crunkhorn

Nature Reviews Drug Discovery (2022)

This article was downloaded by **calibre** from <https://www.nature.com/articles/s41586-022-04654-9>

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

[Download PDF](#)

- Article
- Open Access
- [Published: 27 April 2022](#)

USP14-regulated allostery of the human proteasome by time-resolved cryo-EM

- [Shuwen Zhang](#) ORCID: [orcid.org/0000-0002-5352-409X](#)^{1,2 na1},
- [Shitao Zou](#)^{1,2 na1},
- [Deyao Yin](#)^{1,2},
- [Lihong Zhao](#)^{1,2},
- [Daniel Finley](#) ORCID: [orcid.org/0000-0003-2076-5863](#)³,
- [Zhaolong Wu](#) ORCID: [orcid.org/0000-0002-0537-6424](#)^{1,2,4} &
- [Youdong Mao](#) ORCID: [orcid.org/0000-0001-9302-2257](#)^{1,2,4,5}

[Nature](#) volume 605, pages 567–574 (2022)

- 13k Accesses
- 72 Altmetric
- [Metrics details](#)

Subjects

- [Cryoelectron microscopy](#)
- [Deubiquitylating enzymes](#)
- [Proteasome](#)

Abstract

Proteasomal degradation of ubiquitylated proteins is tightly regulated at multiple levels^{1,2,3}. A primary regulatory checkpoint is the removal of ubiquitin chains from substrates by the deubiquitylating enzyme ubiquitin-specific protease 14 (USP14), which reversibly binds the proteasome and confers the ability to edit and reject substrates. How USP14 is activated and regulates proteasome function remain unknown^{4,5,6,7}. Here we present high-resolution cryo-electron microscopy structures of human USP14 in complex with the 26S proteasome in 13 distinct conformational states captured during degradation of polyubiquitylated proteins. Time-resolved cryo-electron microscopy analysis of the conformational continuum revealed two parallel pathways of proteasome state transitions induced by USP14, and captured transient conversion of substrate-engaged intermediates into substrate-inhibited intermediates. On the substrate-engaged pathway, ubiquitin-dependent activation of USP14 allosterically reprograms the conformational landscape of the AAA-ATPase motor and stimulates opening of the core particle gate^{8,9,10}, enabling observation of a near-complete cycle of asymmetric ATP hydrolysis around the ATPase ring during processive substrate unfolding. Dynamic USP14–ATPase interactions decouple the ATPase activity from RPN11-catalysed deubiquitylation^{11,12,13} and kinetically introduce three regulatory checkpoints on the proteasome, at the steps of ubiquitin recognition, substrate translocation initiation and ubiquitin chain recycling. These findings provide insights into the complete functional cycle of the USP14-regulated proteasome and establish mechanistic foundations for the discovery of USP14-targeted therapies.

Main

The majority of cellular proteins in eukaryotes are targeted to the 26S proteasome for degradation by ubiquitylation pathways, which regulate major aspects of cellular processes^{1,10}. The proteasome holoenzyme is assembled from a cylindrical 20S core particle (CP) capped with one or two 19S regulatory particles (RPs), each consisting of the lid and base subcomplexes^{1,10}. The ring-like heterohexameric motor of the AAA (ATPase associated with diverse cellular activities) family of adenosine

triphosphatase (ATPase) in the base regulates substrate processing in the proteasome via multiple modes of coordinated ATP hydrolysis¹⁰. The proteasome is dynamically regulated by numerous proteins that reversibly associate with it via mechanisms that remain unknown³.

USP14 is one of the three proteasome-associated deubiquitinating enzymes² (DUBs) and is crucially involved in the regulation of proteostasis, inflammation, neurodegeneration, tumorigenesis and viral infection^{2,14,15}. USP14 is a potential therapeutic target for treating cancer, inflammatory and neurodegenerative diseases^{6,14,16,17}. In common with its yeast orthologue Ubp6, USP14 has major roles in proteasome regulation and is prominently activated upon reversible association with the proteasome^{4,5,6,7}. USP14 and Ubp6 stabilize many cellular proteins against proteasomal degradation by ubiquitin chain disassembly as well as noncatalytically^{18,19,20}. Unlike the stoichiometric subunit RPN11 (also known as PSMD14), the DUB activity of which is coupled to ATP-driven substrate translocation, USP14 catalyses removal of supernumerary ubiquitin chains on a substrate en bloc, independently of ATPase activity, until a single chain remains⁷. Paradoxically, binding of ubiquitylated substrates to USP14 stimulates proteasomal ATPase activity and CP gate opening^{20,21,22,23}. The molecular mechanisms underlying USP14 activation and its regulation of the proteasome remain unknown.

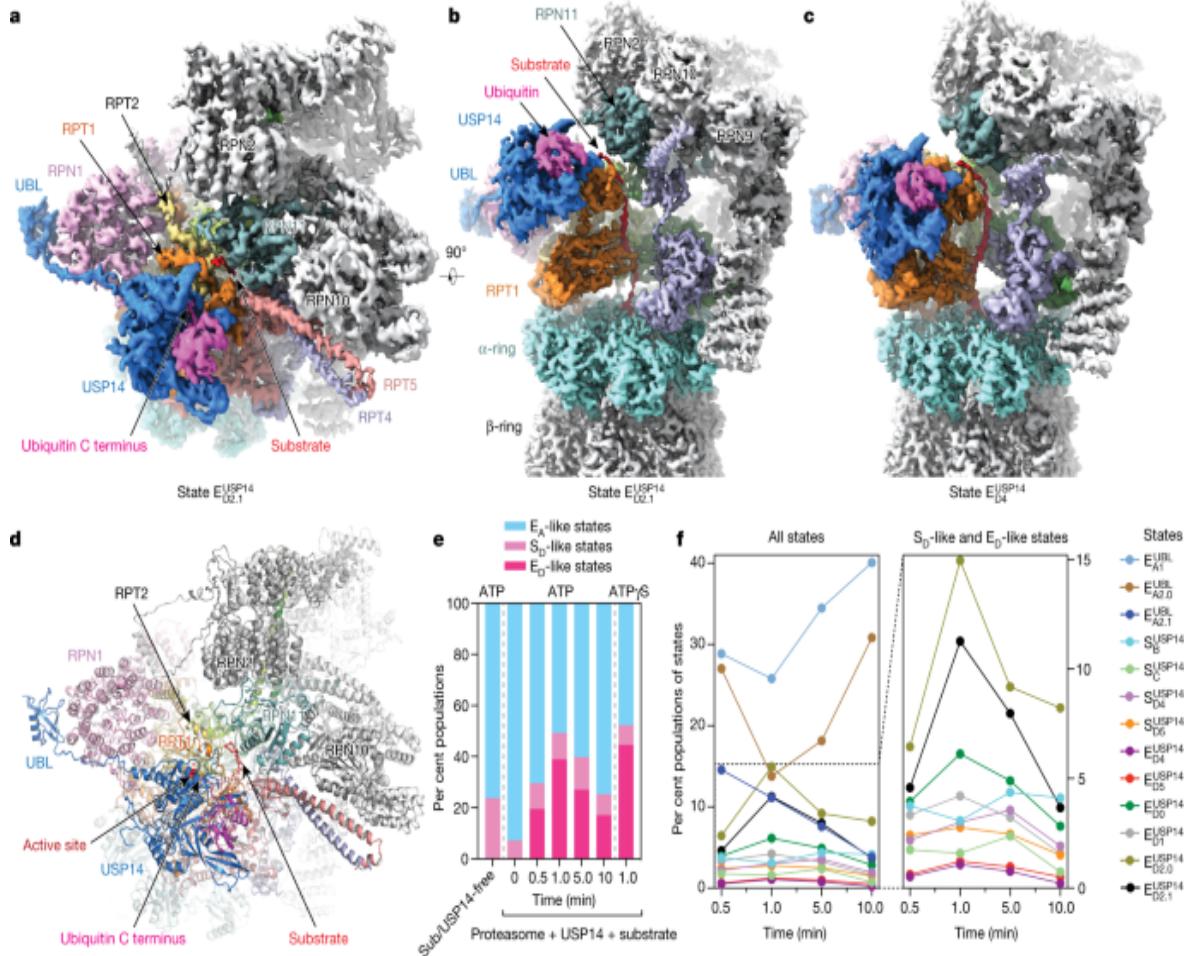
Previous cryo-electron microscopy (cryo-EM) studies have determined the atomic structures of the USP14-free, substrate-engaged proteasomes at key functional steps, including ubiquitin recognition (states E_{A1} and E_{A2}), deubiquitylation (state E_B), translocation initiation (states E_{C1} and E_{C2}) and processive degradation (states E_{D1} and E_{D2})¹⁰. Early cryo-EM reconstructions have revealed the approximate locations of USP14 and Ubp6 in the proteasome^{23,24,25,26}. However, the insufficient resolution and the absence of polyubiquitylated substrates in these studies preclude understanding of USP14-mediated proteasome regulation at the atomic level. In addition, the highly dynamic nature of USP14–proteasome association has hampered structural determination of full-length USP14. Here we report time-resolved cryo-EM studies of human USP14 in complex with functional proteasome in the act of substrate degradation. Our structural

and functional analyses portray a dynamic picture of USP14–proteasome interactions and reveal the mechanism of allosteric ‘tug-of-war’ between USP14 and the proteasome for deciding substrate fate.

Visualizing intermediates of USP14–proteasome

To prepare a substrate-engaged USP14–proteasome complex, we separately purified human USP14, RPN13 (also known as ADRM1) and USP14-free 26S proteasome (Extended Data Fig. 1). Although the ubiquitin receptor RPN13 appears to be present in the purified USP14-free proteasome (Extended Data Fig. 1e), it was absent from previous cryo-EM reconstructions of human proteasomes^{8,9,10,25,27,28,29}, probably due to its sub-stoichiometric levels. In an attempt to saturate the proteasome with RPN13, and thereby enhance substrate recruitment by the USP14-bound proteasome, we mixed stoichiometric excesses of both purified USP14 and RPN13 with the USP14-free proteasome before substrate addition. We used Sic1^{PY} conjugated with Lys63-linked polyubiquitin chains ($\text{Ub}_n\text{--Sic1}^{\text{PY}}$) as a model substrate⁷. Lys63-linked chains are the second most abundant ubiquitin linkages in mammals and regulate essential intracellular functions such as endocytosis, DNA repair and immune responses¹. We collected cryo-EM datasets for samples prepared by cryo-plunging at 0.5, 1, 5 and 10 min after mixing $\text{Ub}_n\text{--Sic1}^{\text{PY}}$ with the USP14-bound proteasome in the presence of 1 mM ATP at 10 °C. Because 3D classification of cryo-EM data indicated that intermediate states were maximized approximately 1 min after substrate addition, we collected considerably more data at this time point (Fig. 1e,f). To facilitate structural determination at high resolution, we collected another large dataset by exchanging ATP with the slowly hydrolysed ATPγS 1 min after substrate addition at 10 °C, which is expected to stall all coexisting intermediate conformations¹⁰. Deep learning-enhanced 3D classification²⁷ enabled us to determine 13 distinct conformational states of the USP14-bound proteasome, including 3 E_A-like, 4 S_D-like and 6 E_D-like conformers, at nominal resolutions of 3.0–3.6 Å (Fig. 1a–d, Extended Data Figs. 2–4, Extended Data Table 1). As expected, each of the 13 states was consistently observed under both the ATP-only and ATP-to-ATPγS exchange conditions despite the differences in state populations (Extended Data Figs. 1r, 2, 5a).

Fig. 1: Time-resolved cryo-EM analysis of the conformational landscape of USP14–proteasome complexes in the act of substrate degradation.



a, b, Cryo-EM density map of the substrate-engaged USP14–proteasome complex in state $\backslash\{\{\backslash\text{rm}\{E\}\}\}_{-}\{\{\backslash\text{rm}\{D2\}\}.1\}^{\{\{\backslash\text{rm}\{USP14\}\}\}}\}$, viewed from the top (**a**) and side (**b**). **c**, Side view of the cryo-EM density map of the substrate-engaged USP14–proteasome complex in state $\backslash\{\{\backslash\text{rm}\{E\}\}\}_{-}\{\{\backslash\text{rm}\{D4\}\}\}^{\{\{\backslash\text{rm}\{USP14\}\}\}}$. Compared to the view of $\backslash\{\{\backslash\text{rm}\{E\}\}\}_{-}\{\{\backslash\text{rm}\{D2\}\}.1\}^{\{\{\backslash\text{rm}\{USP14\}\}\}}$ in **b**, USP14 is rotated about 30° to dock onto the AAA domain of RPT1. To visualize the substrate density inside the AAA-ATPase motor, the density of RPT5 is omitted in both **b** and **c**. **d**, Atomic model of state $\backslash\{\{\backslash\text{rm}\{E\}\}\}_{-}\{\{\backslash\text{rm}\{D2\}\}.1\}^{\{\{\backslash\text{rm}\{USP14\}\}\}}$ viewed from the same perspective as in **a**. **e**, Kinetic changes of overall particle populations of S_D -like and E_D -like states versus E_A -like states obtained from time-resolved cryo-EM analysis. E_A -like states include \backslash

$\{\{\backslash rm\{E\}\}\}_\{\{\backslash rm\{A1\}\}\}^\{\{\backslash rm\{UBL\}\}\}\}, \backslash$
 $\{\{\backslash rm\{E\}\}\}_\{\{\backslash rm\{A2\}\}.0\}^\{\{\backslash rm\{UBL\}\}\}\}$ and
 $\{\{\backslash rm\{E\}\}\}_\{\{\backslash rm\{A2\}\}.1\}^\{\{\backslash rm\{UBL\}\}\}\}$). S_D -like states include
 $\{\{\backslash rm\{S\}\}\}_\{\{\backslash rm\{B\}\}\}^\{\{\backslash rm\{USP14\}\}\}\}, \backslash$
 $\{\{\backslash rm\{S\}\}\}_\{\{\backslash rm\{C\}\}\}^\{\{\backslash rm\{USP14\}\}\}\}\}$ and
 $\{\{\backslash rm\{S\}\}\}_\{\{\backslash rm\{D4\}\}\}^\{\{\backslash rm\{USP14\}\}\}\}$ and
 $\{\{\backslash rm\{S\}\}\}_\{\{\backslash rm\{D5\}\}\}^\{\{\backslash rm\{USP14\}\}\}\}$). E_D -like states include
 $\{\{\backslash rm\{E\}\}\}_\{\{\backslash rm\{D4\}\}\}^\{\{\backslash rm\{USP14\}\}\}\}, \backslash$
 $\{\{\backslash rm\{E\}\}\}_\{\{\backslash rm\{D5\}\}\}^\{\{\backslash rm\{USP14\}\}\}\}, \backslash$
 $\{\{\backslash rm\{E\}\}\}_\{\{\backslash rm\{D0\}\}\}^\{\{\backslash rm\{USP14\}\}\}\}, \backslash$
 $\{\{\backslash rm\{E\}\}\}_\{\{\backslash rm\{D1\}\}\}^\{\{\backslash rm\{USP14\}\}\}\}, \backslash$
 $\{\{\backslash rm\{E\}\}\}_\{\{\backslash rm\{D2\}\}.0\}^\{\{\backslash rm\{USP14\}\}\}\}$ and
 $\{\{\backslash rm\{E\}\}\}_\{\{\backslash rm\{D2\}\}.1\}^\{\{\backslash rm\{USP14\}\}\}\}$). The control consists of previously reported data for substrate-free, USP14-free proteasome⁸. f, Kinetic changes of the particle populations of 13 coexisting conformational states of USP14-bound proteasome from the cryo-EM samples made at different time points after mixing the substrate with the USP14–proteasome complex in the presence of 1 mM ATP at 10 °C. Three substrate-inhibited intermediates ($\{\{\backslash rm\{S\}\}\}_\{\{\backslash rm\{B\}\}\}^\{\{\backslash rm\{USP14\}\}\}\}$, $\{\{\backslash rm\{S\}\}\}_\{\{\backslash rm\{C\}\}\}^\{\{\backslash rm\{USP14\}\}\}\}$ and $\{\{\backslash rm\{S\}\}\}_\{\{\backslash rm\{D4\}\}\}^\{\{\backslash rm\{USP14\}\}\}\}$) reach their maximal populations at around 5 min, in contrast to state
 $\{\{\backslash rm\{S\}\}\}_\{\{\backslash rm\{D5\}\}\}^\{\{\backslash rm\{USP14\}\}\}\}$ and six substrate-engaged states, which all reach their maximal populations at approximately 1 min. The number of particles used in e and f are provided in Extended Data Fig. 2b, c.

The six E_D -like substrate-engaged states—designated
 $\{\{\backslash rm\{E\}\}\}_\{\{\backslash rm\{D0\}\}\}^\{\{\backslash rm\{USP14\}\}\}\}, \backslash$
 $\{\{\backslash rm\{E\}\}\}_\{\{\backslash rm\{D1\}\}\}^\{\{\backslash rm\{USP14\}\}\}\}, \backslash$
 $\{\{\backslash rm\{E\}\}\}_\{\{\backslash rm\{D2\}\}.0\}^\{\{\backslash rm\{USP14\}\}\}\}, \backslash$
 $\{\{\backslash rm\{E\}\}\}_\{\{\backslash rm\{D2\}\}.1\}^\{\{\backslash rm\{USP14\}\}\}\}, \backslash$
 $\{\{\backslash rm\{E\}\}\}_\{\{\backslash rm\{D4\}\}\}^\{\{\backslash rm\{USP14\}\}\}\}$ and
 $\{\{\backslash rm\{E\}\}\}_\{\{\backslash rm\{D5\}\}\}^\{\{\backslash rm\{USP14\}\}\}\}$ —captured sequential intermediates during processive substrate unfolding and translocation, which are compatible with the hand-over-hand translocation model^{10,30} (Extended Data Figs. 6, 7, Extended Data Table 2). The cryo-EM density of USP14 in

state $\{\{ \text{rm} \{ E \} \} _ \{\{ \text{rm} \{ D2 \} \} .1\} ^{\{\{ \text{rm} \{ USP14 \} \}\}}\}$ is of sufficient quality to allow atomic modelling of full-length USP14 (Extended Data Fig. 4c, d). An unfolded polypeptide substrate is observed in the AAA-ATPase channel in all of the E_D -like states. These states also exhibit an open CP gate, with five C-terminal tails (C-tails) of ATPase subunits (RPT1–RPT6, excluding RPT4) inserted into the inter-subunit surface pockets on the α -ring (α -pockets) of the CP^{8,9,10,30,31} (Fig. 3a, Extended Data Fig. 6h).

There are no substrate densities observed in the AAA-ATPase channel of the four S_D -like states, designated \

$\{\{ \text{rm} \{ S \} \} _ \{\{ \text{rm} \{ B \} \} ^{\{\{ \text{rm} \{ USP14 \} \}\}}\}, \$
 $\{\{ \text{rm} \{ S \} \} _ \{\{ \text{rm} \{ C \} \} ^{\{\{ \text{rm} \{ USP14 \} \}\}}\}, \$
 $\{\{ \text{rm} \{ S \} \} _ \{\{ \text{rm} \{ D4 \} \} ^{\{\{ \text{rm} \{ USP14 \} \}\}}\} \text{ and } \$
 $\{\{ \text{rm} \{ S \} \} _ \{\{ \text{rm} \{ D5 \} \} ^{\{\{ \text{rm} \{ USP14 \} \}\}}\}.$ Substrate insertion into the AAA-ATPase channel is sterically inhibited in these states, because RPN11 blocks the substrate entrance at the oligonucleotide- or oligosaccharide-binding (OB) ring of the AAA-ATPase motor (Extended Data Fig. 7f). The AAA-ATPase conformations of \

$\{\{ \text{rm} \{ S \} \} _ \{\{ \text{rm} \{ B \} \} ^{\{\{ \text{rm} \{ USP14 \} \}\}}\}, \$
 $\{\{ \text{rm} \{ S \} \} _ \{\{ \text{rm} \{ C \} \} ^{\{\{ \text{rm} \{ USP14 \} \}\}}\}, \$
 $\{\{ \text{rm} \{ S \} \} _ \{\{ \text{rm} \{ D4 \} \} ^{\{\{ \text{rm} \{ USP14 \} \}\}}\} \text{ and } \$
 $\{\{ \text{rm} \{ S \} \} _ \{\{ \text{rm} \{ D5 \} \} ^{\{\{ \text{rm} \{ USP14 \} \}\}}\} \text{ resemble those of } S_B, S_{D1}, \$
 $\{\{ \text{rm} \{ E \} \} _ \{\{ \text{rm} \{ D4 \} \} ^{\{\{ \text{rm} \{ USP14 \} \}\}}\} \text{ and } \$
 $\{\{ \text{rm} \{ E \} \} _ \{\{ \text{rm} \{ D5 \} \} ^{\{\{ \text{rm} \{ USP14 \} \}\}}\}, \text{ respectively}^{8,9} \text{ (Extended Data Figs. 5f, 7d). However, the ATPase-CP interfaces of states } \$
 $\{\{ \text{rm} \{ S \} \} _ \{\{ \text{rm} \{ B \} \} ^{\{\{ \text{rm} \{ USP14 \} \}\}}\}, \$
 $\{\{ \text{rm} \{ S \} \} _ \{\{ \text{rm} \{ C \} \} ^{\{\{ \text{rm} \{ USP14 \} \}\}}\} \text{ and } \$
 $\{\{ \text{rm} \{ S \} \} _ \{\{ \text{rm} \{ D \} \} 4/5\} ^{\{\{ \text{rm} \{ U \} \} \{\text{rm} \{ S \} \} \{\text{rm} \{ P \} \} 14\}}$ resemble those of states E_A, E_C and E_D , where two, four and five RPT C-tails are inserted into the α -pockets, respectively¹⁰ (Extended Data Fig. 6h). Thus, the CP gate remains closed in $\{\{ \text{rm} \{ S \} \} _ \{\{ \text{rm} \{ B \} \} ^{\{\{ \text{rm} \{ USP14 \} \}\}}\}$ and $\{\{ \text{rm} \{ S \} \} _ \{\{ \text{rm} \{ C \} \} ^{\{\{ \text{rm} \{ USP14 \} \}\}}\}$, but is open in $\{\{ \text{rm} \{ S \} \} _ \{\{ \text{rm} \{ D4 \} \} ^{\{\{ \text{rm} \{ USP14 \} \}\}}\}$ and $\{\{ \text{rm} \{ S \} \} _ \{\{ \text{rm} \{ D5 \} \} ^{\{\{ \text{rm} \{ USP14 \} \}\}}\}.$

Human USP14 comprises 494 amino acids and features a 9-kDa ubiquitin-like (UBL) domain at its N terminus, followed by a 43-kDa USP domain

joined via a flexible linker region of 23 amino acids. The overall USP14 structure bridges the RPN1 and RPT1 subunits in all the S_D-like and E_D-like states, in agreement with previous low-resolution studies^{23,24,25,26} (Extended Data Fig. 5c). By contrast, the three E_A-like states, designated \

\{\{\rm{E}\}\}_{{\{\rm{A}1\}}}\}^{{\{\rm{UBL}\}}}\}), \\\{\{\rm{E}\}\}_{{\{\rm{A}2\}}}.0\}^{{\{\rm{UBL}\}}}) and \\\{\{\rm{E}\}\}_{{\{\rm{A}2\}}}.1\}^{{\{\rm{UBL}\}}}\}), show no visible density for the catalytic USP domain of USP14, although a density consistent with the UBL domain is observed at the T2 site of RPN1^{27,32} (also known as PSMD2) (Extended Data Figs. 4a, 7e).

Time-resolved conformational continuum

Time-dependent quantification of the degraded Ub_n–Sic1^{PY} substrate indicates that USP14 reduced the degradation rate by approximately two-fold compared with that of the USP14-free proteasome⁷ (Extended Data Fig. 1k). The majority (about 80%) of substrates were eventually degraded by the USP14–proteasome mixture after 30 min of reaction under the same conditions used for our time-resolved cryo-EM analysis. Notably, the substrate-engaged and substrate-inhibited intermediates coexisted at all measured time points and reached their maximal per cent levels at approximately 1 and 5 min, respectively, after substrate addition (Fig. 1e). The population of \{\{\rm{S}\}\}_{{\{\rm{B}\}}}\}^{{\{\rm{USP14}\}}}\} became the largest among the substrate-inhibited states in 5–10 min (Fig. 1f). The overall population of substrate-inhibited intermediates varied within a small range of 7.7–10.3% from 0.5 to 5 min and was comparable to the 7.3 % of the S_D-like states before substrate addition (Fig. 1e,f, Extended Data Fig. 2b,c), suggesting that the substrate-inhibited pathway was induced in parallel to the substrate-engaged pathway and was promoted by USP14 despite the stoichiometric excess of ubiquitylated substrates. Consistent with previous studies²³, these results indicate that the substrate-engaged intermediates were converted to substrate-inhibited states upon termination of substrate translocation.

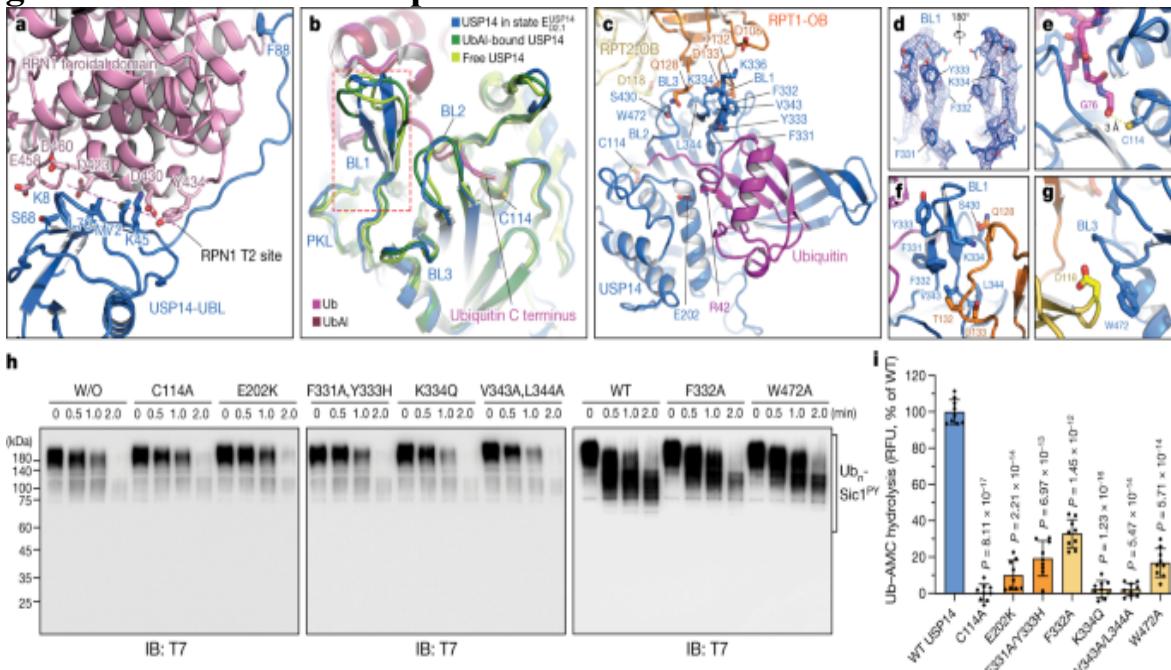
Despite exhaustive 3D classification, we did not observe in any USP14-loaded experimental conditions the proteasome states E_B and E_C, which

represent RPN11-mediated deubiquitylation and translocation initiation prior to the CP gate opening, respectively¹⁰. This implies that USP14 prevents the proteasome from assuming the conformation of RPN11-catalysed deubiquitylation. In contrast to the RPN11-bound ubiquitin in states \(\{\{\backslash rm{E}\}\}\}_{\{\{\backslash rm{A2}\}\}.0}^{\{\{\backslash rm{UBL}\}\}}\})\) and \(\{\{\backslash rm{E}\}\}\}_{\{\{\backslash rm{A2}\}\}.1}^{\{\{\backslash rm{UBL}\}\}}\}), no ubiquitin was observed on RPN11 in any state in which USP14 is engaged with ubiquitin (Fig. 1a–d, Extended Data Figs. 4a, 7a–c), suggesting that activated USP14 and RPN11 do not bind ubiquitin simultaneously.

Dynamic USP14–proteasome interactions

The UBL domain of USP14 binds the RPN1 T2 site via a hydrophobic patch centred on residue Leu70, which is structurally homologous to the Ile44 patch of ubiquitin (Fig. 2a). The RPN1 T2 site is composed of residues Asp423, Leu426, Asp430, Tyr434, Glu458, Asp460 and Leu465 on two adjacent helix-loop regions, in agreement with previous findings^{27,32}. The N-terminal stretch of the linker (residues Ala77 to Phe88) in USP14 appears to bind the ridge of the toroid domain of RPN1 (Fig. 2a).

Fig. 2: Structural basis of proteasome-mediated activation of USP14.



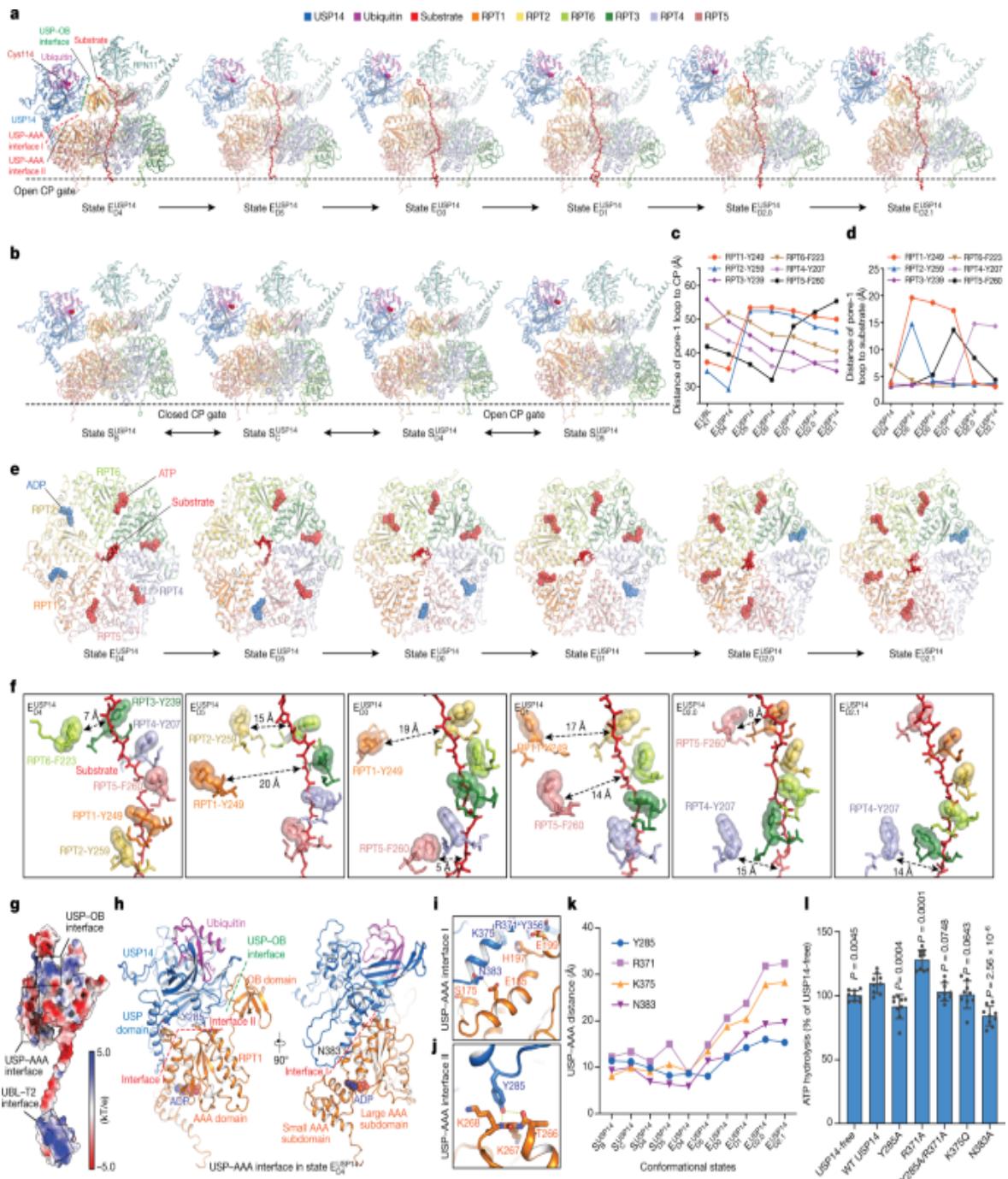
a, Side-chain interactions between the USP14 UBL domain and the RPN1 T2 site in the proteasome state \(\{\{\backslash rm\{E\}\}\}_\{\{\backslash rm\{D2\}\}.1\}^{\{\{\backslash rm\{USP14\}\}\}}\). **b**, Structural comparison of the blocking loops by superimposing the USP14 structure in state \(\{\{\backslash rm\{E\}\}\}_\{\{\backslash rm\{D2\}\}.1\}^{\{\{\backslash rm\{USP14\}\}\}}\) with two crystal structures of USP14 in its isolated form (PDB ID: 2AYN) and in complex with ubiquitin aldehyde⁵ (UbAl) (PDB ID: 2AYO). **c**, Magnified view of the ubiquitin–USP–OB sandwich architecture in state \(\{\{\backslash rm\{E\}\}\}_\{\{\backslash rm\{D2\}\}.1\}^{\{\{\backslash rm\{USP14\}\}\}}\). **d**, Local cryo-EM density of the BL1 motif in state \(\{\{\backslash rm\{E\}\}\}_\{\{\backslash rm\{D2\}\}.1\}^{\{\{\backslash rm\{USP14\}\}\}}\) in mesh representation superimposed with its atomic model in cartoon representation from two opposite orientations, showing its β-hairpin conformation. **e–g**, Magnified views of the interfaces between the catalytic Cys114 of USP14 and the C-terminal Gly76 of ubiquitin (**e**), between the USP14 BL1 motif and the RPT1 OB domain (**f**), and between the USP14 BL3 motif and the RPT2 OB domain (**g**). Key residues mediating the inter-molecular interactions are shown in stick representation in **a–g**. **h**, In vitro degradation of Ub_n–Sic1^{PY} by the human proteasome assembled with USP14 variants at 37 °C, analysed by SDS–PAGE and western blot using anti-T7 antibody to visualize the fusion protein T7–Sic1^{PY}. See Supplementary Fig. 1 for gel source data. These experiments were repeated independently three times with consistent results. The proteasome without USP14 (no USP14; labelled W/O above the leftmost lanes) is used as a negative control. Lanes labelled WT correspond to the proteasome bound to wild-type USP14. **i**, Ubiquitin–AMC hydrolysis by the USP14 mutants dictates their DUB activity in the proteasome. RFU, relative fluorescence units. RFU values at 60 min are shown. All labelled *P* values were computed against the wild-type USP14 using a two-tailed unpaired *t*-test. Data are mean ± s.d. from three independent experiments. Each experiment includes three replicates. The quantification of wild-type USP14 was used as a denominator to normalize all measurements.

Source data

USP14 interacts with both the OB and AAA domains of the ATPase ring. The USP–OB interaction is mediated by the blocking loops 1 (BL1), 2 (BL2) and 3 (BL3) of the USP domain (Fig. 2b), which buries a solvent-

accessible area of approximately 527 \AA^2 on USP14. BL1 makes the most extensive interface with the OB ring around residues Gln128 and Asp133 of RPT1 (also known as PSMC2), whereas Ser430 in BL2 and Trp472 in BL3 interact with Gln128 of RPT1 and Asp118 of RPT2 (also known as PSMC1), respectively (Fig. 2c). The movement of the USP domain between different proteasome states appears to adapt to the rocking motion of the AAA-ATPase motor and maintains its interaction with the OB ring (Fig. 3a–d, Extended Data Fig. 8a).

Fig. 3: Structural dynamics and mechanism of allosteric regulation of the AAA-ATPase motor by USP14.



a, b, Side-by-side comparison of the USP14–ATPase subcomplex structures aligned against the CP in six substrate-engaged states (**a**) and four substrate-inhibited states (**b**). **c, d**, Plots of distance from the pore-1 loop of each ATPase to the CP (**c**) and to the substrate (**d**) in distinct states. **e**, The AAA domain structures of the ATPase motor in six substrate-engaged states. **f**, Varying architecture of the pore-1 loop staircase interacting with the substrate.

substrate in distinct states. The distances from disengaged pore-1 loops to the substrate are marked. The side chains of the pore-1 loop residues, featuring a consensus sequence of K/M-Y/F-V/L/I, are shown in stick representation, with the aromatic residues highlighted in transparent sphere representation. **g**, Electrostatic surface representation of the full-length USP14, coloured according to electrostatic potential from red (-5.0 kT e^{-1} , negatively charged) to blue (5.0 kT e^{-1} , positively charged). **h**, Atomic model of the USP14–RPT1 subcomplex in state $\backslash\{\{\backslash\text{rm}\{E\}\}\}_\backslash\{\{\backslash\text{rm}\{D4\}\}\}^\backslash\{\{\backslash\text{rm}\{USP14\}\}\}\backslash$ in cartoon representation. **i**, **j**, Magnified views of the USP–AAA interface I (**i**) and II (**j**) in state $\backslash\{\{\backslash\text{rm}\{E\}\}\}_\backslash\{\{\backslash\text{rm}\{D4\}\}\}^\backslash\{\{\backslash\text{rm}\{USP14\}\}\}\backslash$, with the interacting pairs of residues in stick representation. **k**, Changes of the USP–AAA interface in distinct states, characterized by measuring the shortest distance between four USP14 residues (Y285, R371, K375 and N383) and the main chains of RPT1 AAA domain. **l**, ATPase activity was quantified by measuring the release of phosphate from ATP hydrolysis of the proteasome. All labelled P values were computed by comparison with wild-type USP14 using a two-tailed unpaired t -test. Data are mean \pm s.d. from three independent experiments, each with three replicates. The quantification of USP14-free proteasome was used as a denominator to normalize all measurements in each experiment.

Source data

In contrast to the USP–OB interface, which is nearly invariant in all S_D -like and E_D -like states, USP–AAA interactions vary prominently (Fig. [3a, b](#)). In states $\backslash\{\{\backslash\text{rm}\{E\}\}\}_\backslash\{\{\backslash\text{rm}\{D0\}\}\}^\backslash\{\{\backslash\text{rm}\{USP14\}\}\}\backslash$, $\backslash\{\{\backslash\text{rm}\{E\}\}\}_\backslash\{\{\backslash\text{rm}\{D1\}\}\}^\backslash\{\{\backslash\text{rm}\{USP14\}\}\}\backslash$, $\backslash\{\{\backslash\text{rm}\{E\}\}\}_\backslash\{\{\backslash\text{rm}\{D2\}\}.0\}^\backslash\{\{\backslash\text{rm}\{USP14\}\}\}\backslash$ and $\backslash\{\{\backslash\text{rm}\{E\}\}\}_\backslash\{\{\backslash\text{rm}\{D2\}\}.1\}^\backslash\{\{\backslash\text{rm}\{USP14\}\}\}\backslash$, the USP domain is flipped up and completely detached from the AAA domain of RPT1. By contrast, the USP domain exhibits differential interactions with the AAA domain of RPT1 in other states (Fig. [3h](#), Extended Data Fig. [8a](#)), suggesting that USP14 preferentially recognizes these AAA conformations and stabilizes these states by direct interactions. The most extensive USP–AAA interactions are observed in state $\backslash\{\{\backslash\text{rm}\{E\}\}\}_\backslash\{\{\backslash\text{rm}\{D4\}\}\}^\backslash\{\{\backslash\text{rm}\{USP14\}\}\}\backslash$ (Fig. [3h–k](#)). In this

conformation, a helix-loop region (residues 371-391) protruding from the USP domain contacts the AAA domain of RPT1, which buries a solvent-accessible area of approximately 850 Å² on USP14.

USP14 activation by the proteasome

Occlusion of the ubiquitin C terminus-binding groove in USP14 by BL1 and BL2 auto-inhibits the DUB activity in the absence of the proteasome⁵.

Comparison of the USP14 structure in state \

({{\rm{E}}})_{{{\rm{D2}}}.1}^{{{\rm{USP14}}}}\) with the crystal structures of the USP domain in isolated and ubiquitin aldehyde-bound forms⁵ reveals differential conformational changes in BL1 and BL2 (Fig. 2b). The BL1 region is an open loop in the crystal structures but is folded into a β-hairpin sandwiched between the OB ring and ubiquitin in the proteasome (Fig. 2b, d). By contrast, the BL2 loop is moved approximately 4 Å to make way for docking of the ubiquitin C terminus into the active site groove. Stabilized by the USP–OB interface, BL1 and BL2 together hold the ubiquitin C terminus in a β-strand conformation (Fig. 2c), in which the main chain carbon of ubiquitin C-terminal Gly76 is placed approximately 3.4 Å away from the sulfur atom of the catalytic Cys114 and is fully detached from the substrate, indicating that the structure represents a post-deubiquitylation state of USP14 (Fig. 2e). This ternary architecture of the ubiquitin–USP–OB sandwich indicates that ubiquitin binding stabilizes the BL1 β-hairpin conformation and the USP interaction with the OB ring. Indeed, microscale thermophoresis (MST) measurements show that the dissociation constant (approximately 44 nM) of USP14 with the proteasome in the presence of Ub_n–Sic1^{PY} is approximately half of that (approximately 95 nM) in the absence of ubiquitin conjugates and one third of that (approximately 137 or 135 nM, respectively) of either UBL or USP domain alone (Extended Data Fig. 11–q). Thus, USP14 affinity towards the proteasome reflects interactions of both its UBL and USP domains and is enhanced by Ub_n–Sic1^{PY}, consistent with previous studies³³.

To test the functional importance of the ubiquitin–USP–OB interface in USP14 activation, we performed structure-based site-directed mutagenesis (Extended Data Fig. 9). Indeed, the USP mutations, including the single

mutant K334Q and the double mutant V343A/L344A at the BL1–OB interface (Fig. 2f), F331A/Y333A at the ubiquitin–BL1 interface, as well as the single mutant E202K that disrupts the salt bridge between Glu202 of USP and Arg42 of ubiquitin (Fig. 2c), all abrogated the DUB activity of USP14—similar to the effect by C114A mutation that removes the active site of USP14—and showed no obvious inhibition of proteasomal degradation (Fig. 2h,i, Extended Data Fig. 9e,f). By contrast, the single mutants BL1(F332A) and BL3(W472A) retained a reduced DUB activity and suppressed proteasome function to a lesser extent (Fig. 2f–i). These contrasting phenotypes substantiate our structural finding that the ubiquitin–USP–OB interactions are essential for USP14 activation towards efficacious deubiquitylation.

The overall structure of full-length USP14 exhibits three major states (Extended Data Fig. 8b), far fewer than the number of USP14-bound proteasome states, suggesting that the conformational entropy of the linker region is greatly reduced upon USP14 assembly and activation on the proteasome. In the linker region, the residues Glu90 and Asp91 are potentially involved in transient interactions with the RPN1 toroid (Extended Data Figs. 7g, 9a). Deletion of residues 93–96 and insertion of TEEQ after residue 92 in the linker suppressed proteasomal degradation of $\text{Ub}_n\text{-Sic1}^{\text{PY}}$ more potently than the double point mutation E90K/D91A (Extended Data Fig. 9d,g). All three USP14 mutants showed 20–30% reductions in DUB activity relative to the wild-type USP14. These observations suggest that the linker length and composition of USP14 may have been evolutionarily optimized for DUB activity, although the UBL and USP domain architecture appears to be functionally robust against variations of the linker region.

Allosteric regulation of ATPase activity

USP14 interacts with the AAA domain of RPT1 via two discrete interfaces in state \(\{\{\text{E}\}\}_\{\{\text{D4}\}\}\^{\{\{\text{USP14}\}\}}\}\) (Fig. 3a, Extended Data Fig. 4f). The primary USP–AAA interface is composed of a negatively charged surface on a helix (residues 371–383) and a loop (residues 384–391) protruding out of the USP domain, which buries the majority of the USP–AAA interface (Fig. 3g,h). The helix region appears to pack against a

convex surface of the large AAA subdomains of RPT1, on which Arg371, Lys375 and Asn383 of USP14 interact with Glu185, His197 and Glu199 of RPT1, whereas the loop region (384–391) reaches the small AAA subdomain of RPT1 (Fig. 3*i*). The second USP–AAA interface is centred on Tyr285 in the PKL loop region of USP14 that interacts with Lys267 and Lys268 of RPT1 (Fig. 3*j*).

In state $\{\{ \backslash \text{rm}\{ E \} \} \}_\{\{ \backslash \text{rm}\{ D4 \} \} \}^\{\{ \backslash \text{rm}\{ \text{USP14} \} \} \}$, ADP is bound to the Walker A motif between the large and small AAA subdomain of RPT1, the dihedral angle of which is directly stabilized by the USP–AAA interaction (Fig. 3*h*). In other substrate-engaged states, in which the USP domain is detached from the AAA domain, the nucleotide-binding pocket of RPT1 is either empty or bound to ATP (Fig. 3*e*, Extended Data Fig. 6*j*). Thus, the USP–AAA interaction energetically stabilizes the ADP-bound conformation of RPT1 and USP–AAA dissociation promotes nucleotide exchange in RPT1. We postulate that the USP–AAA interaction may allosterically promote the ATPase activity during substrate degradation. Indeed, wild-type USP14 enhanced the ATPase activity during proteasomal degradation of $\text{Ub}_n\text{--Sic1}^{\text{PY}}$ by approximately 10% relative to that of the USP14-free proteasome (Fig. 3*l*).

To test the functional roles of the USP–AAA interfaces, we mutated several USP14 residues at the USP–AAA interfaces to alanine. Although none of these USP14 mutants exhibited any observable defects in the DUB activity of USP14 or the peptidase activity of USP14-bound proteasome (Extended Data Fig. 9*c,h*), they perturbed the ATPase activity in the presence of polyubiquitylated substrate, in line with previous studies^{21,22,23,34,35}. Two mutants, Y285A and N383A, showed approximately 20–30% reduction of the ATPase activity to a level 10–20% below that of the USP14-free proteasome (Fig. 3*l*). As Tyr285 is located halfway between the primary USP–AAA and USP–OB interfaces (Fig. 3*f,h*), it may be mechanically important in transmitting the allosteric effect from USP to AAA. Both Y285A and N383A mutations potentially reduce the overall USP association with the large AAA subdomain of RPT1 (Fig. 3*i*). By contrast, the single mutant R371A enhanced the ATPase activity by approximately 20% relative to the wild-type USP14, probably by tightening the USP–AAA interactions; whereas K375Q showed no significant effect on the ATPase activity,

probably because only the main chain of K375 interacts with RPT1 His197 (Fig. 3j). Consistently, the double mutant Y285A/R371A restored the ATPase rate to the approximate level of the USP14-free proteasome, presumably owing to the cancellation of two counteracting allosteric effects between the two mutated residues (Fig. 3l). Altogether, our structure-guided mutagenesis supports the notion that the USP–AAA interfaces mediate the non-catalytic, allosteric regulation of the ATPase activity in the proteasome.

Asymmetric ATP hydrolysis around ATPase ring

The six substrate-engaged USP14–proteasome structures characterize a detailed intermediate sequence of substrate translocation, which deviates substantially from the pathway of translocation initiation of the USP14-free proteasome¹⁰ (Extended Data Figs. 5, 6). Whereas the RP conformations of states $\{\{ \rm{E} \}\}_{\{\{ \rm{D1} \}\}}^{\{\{ \rm{USP14} \}\}}$ and $\{\{ \rm{E} \}\}_{\{\{ \rm{D2} \}\}.1}^{\{\{ \rm{USP14} \}\}}$ closely resemble those of USP14-free proteasomes in previously reported states E_{D1} and E_{D2} , respectively, other substrate-engaged states of USP14–proteasome present distinct, hitherto unknown RP conformations¹⁰ (Extended Data Figs. 5d–f, 6). Of note, the AAA-ATPase conformations in states $\{\{ \rm{E} \}\}_{\{\{ \rm{D4} \}\}}^{\{\{ \rm{USP14} \}\}}$ and $\{\{ \rm{E} \}\}_{\{\{ \rm{D5} \}\}}^{\{\{ \rm{USP14} \}\}}$ can be compared to those of USP14-free, closed-CP states E_B and E_{C1} , respectively, albeit with defined structural differences^{10,27} (Extended Data Fig. 5d, f). In particular, the pore-1 loop of RPT6 in state $\{\{ \rm{E} \}\}_{\{\{ \rm{D4} \}\}}^{\{\{ \rm{USP14} \}\}}$ has considerably moved up towards the substrate as compared with that in state E_B (Extended Data Fig. 5e). Given the absence of the E_B - and E_C -like states in the presence of USP14, our structural data together prompt the hypothesis that $\{\{ \rm{E} \}\}_{\{\{ \rm{D4} \}\}}^{\{\{ \rm{USP14} \}\}}$ and $\{\{ \rm{E} \}\}_{\{\{ \rm{D5} \}\}}^{\{\{ \rm{USP14} \}\}}$ may replace the USP14-free states E_B and E_C during initiation of substrate translocation. Thus, we infer that these states present a continuum of USP14-altered conformations following the state-transition sequence $\{\{ \rm{E} \}\}_{\{\{ \rm{A1} \}\}}^{\{\{ \rm{UBL} \}\}} \rightarrow \{\{ \rm{E} \}\}_{\{\{ \rm{D4} \}\}}^{\{\{ \rm{USP14} \}\}}$.

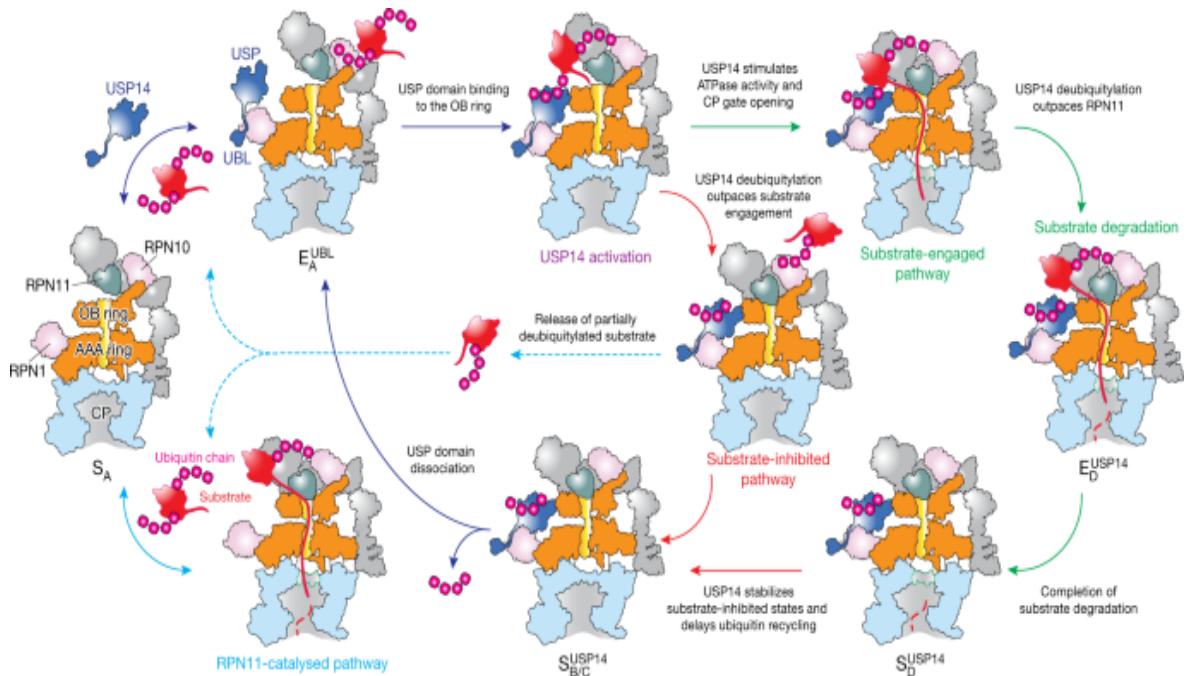
$\{\{\backslash\text{rm}\{E\}\}\}_{-}\{\{\backslash\text{rm}\{D5\}\}\}^{\wedge}\{\{\backslash\text{rm}\{USP14\}\}\}\rightarrow\backslash$
 $\{\{\backslash\text{rm}\{E\}\}\}_{-}\{\{\backslash\text{rm}\{D0\}\}\}^{\wedge}\{\{\backslash\text{rm}\{USP14\}\}\}\rightarrow\backslash$
 $\{\{\backslash\text{rm}\{E\}\}\}_{-}\{\{\backslash\text{rm}\{D1\}\}\}^{\wedge}\{\{\backslash\text{rm}\{USP14\}\}\}\rightarrow\backslash$
 $\{\{\backslash\text{rm}\{E\}\}\}_{-}\{\{\backslash\text{rm}\{D2\}\}\}.0\}^{\wedge}\{\{\backslash\text{rm}\{USP14\}\}\}\rightarrow\backslash$
 $\{\{\backslash\text{rm}\{E\}\}\}_{-}\{\{\backslash\text{rm}\{D2\}\}\}.1\}^{\wedge}\{\{\backslash\text{rm}\{USP14\}\}\}\)$ (Fig. 3c, Extended Data Fig. 7i,j). In support of this sequence assignment, the RPT3 pore-1 loop is moved sequentially from the top to the bottom of the substrate–pore loop staircase, whereas coordinated ATP hydrolysis navigates around the ATPase ring for a near-complete cycle (Fig. 3c–f, Extended Data Fig. 6a–f).

The newly resolved states fill a major gap in our understanding of substrate translocation by the proteasomal AAA-ATPase motor. It has remained unclear whether a strict sequential hand-over-hand mechanism is used by the AAA-ATPase motor for processive substrate translocation³⁶. Of note, four of the six E_D-like states— $\{\{\backslash\text{rm}\{E\}\}\}_{-}\{\{\backslash\text{rm}\{D5\}\}\}^{\wedge}\{\{\backslash\text{rm}\{USP14\}\}\}\backslash$, $\{\{\backslash\text{rm}\{E\}\}\}_{-}\{\{\backslash\text{rm}\{D0\}\}\}^{\wedge}\{\{\backslash\text{rm}\{USP14\}\}\}\backslash$, $\{\{\backslash\text{rm}\{E\}\}\}_{-}\{\{\backslash\text{rm}\{D1\}\}\}^{\wedge}\{\{\backslash\text{rm}\{USP14\}\}\}\backslash$ and $\{\{\backslash\text{rm}\{E\}\}\}_{-}\{\{\backslash\text{rm}\{D2\}\}\}.0\}^{\wedge}\{\{\backslash\text{rm}\{USP14\}\}\}\backslash$ —exhibit an AAA-ATPase conformation with two adjacent RPT pore-1 loops disengaged from the substrate: one moving away from the substrate while releasing ADP for nucleotide exchange, and the other moving up towards the substrate upon binding ATP for substrate re-engagement (Fig. 3d–f). Recent studies of the proteasome have found that substrate re-engagement of an ATPase subunit can be the rate-limiting step in single-nucleotide exchange dynamics²⁷. Coupling of conformational changes between adjacent AAA domains may vary from tightly coupled kinetics—that is, only one dissociated pore-1 loop at a time—to loosely coupled kinetics exhibiting two adjacent pore-1 loops dissociated from the substrate simultaneously^{10,27}. Thus, two adjacent pore-1 loops dissociated from the substrate represent an inevitable intermediate in the sequential hand-over-hand model of substrate translocation, which would be very short-lived in tightly coupled kinetics³⁷. Under the allosteric influence of USP14, such kinetics can become loosely coupled around RPT1, leading to meta-stabilization of these short-lived intermediate states. The observations of both tightly and loosely coupled kinetics at different locations around the ATPase ring rationalize broken symmetry of coordinated ATP hydrolysis in the AAA-ATPase motor—an effect that has been previously informed by functional studies^{38,39}.

Insights into proteasome regulation by USP14

Our structural, kinetic and functional data collectively provide insights into how USP14 regulates proteasome activity at multiple checkpoints by inducing parallel pathways of proteasome state transitions (Fig. 4, Extended Data Fig. 7j). The first checkpoint is at the initial ubiquitin recognition before substrate engagement with the ATPases. USP14 assembly on the proteasome is initiated by its UBL domain binding to the T2 site of the RPN1 toroid^{27,32} as suggested in state \(\{\{\rm{E}\}\}_ \{\{\rm{A1}\}\}^{\{\{\rm{UBL}\}\}}\). Binding of the UBL domain alone has been previously suggested to allosterically stimulate proteasome activity^{34,35}. This is followed by a distinct association step of its USP domain with the OB ring of the AAA-ATPase motor in a ubiquitin-dependent manner, which allosterically activates USP14 and reciprocally stabilizes the proteasome in the substrate-inhibited states. It appears that the USP-ATPase association is allosterically precluded at large once substrate-conjugated ubiquitin is recruited to RPN11, as suggested by states \(\{\{\rm{E}\}\}_ \{\{\rm{A2}\}\}.0\}^{\{\{\rm{UBL}\}\}}\)) and \(\{\{\rm{E}\}\}_ \{\{\rm{A2}\}\}.1\}^{\{\{\rm{UBL}\}\}}\)) (Extended Data Fig. 7a, b). Such a competition in ubiquitin recruitment forces the proteasome to bypass the conformational transition pathway represented by USP14-free states E_B and E_C, which are needed to couple RPN11-catalysed deubiquitylation with ATP-dependent substrate engagement and translocation initiation¹⁰. Thus, the first checkpoint allows USP14 to allosterically compete against RPN11 in accepting substrate-conjugated ubiquitin.

Fig. 4: Proposed model of USP14-mediated regulation of proteasome function.



USP14 binding to the RPN1 and RPT1 subunits of the proteasome primes USP14 activation, whereas ubiquitin–substrate conjugates recruited to the proteasome’s ubiquitin receptors facilitate ubiquitin recognition by USP14. RPN11-catalysed pathway (turquoise solid arrow) is allosterically excluded once USP14 is recruited to the proteasome (dark blue arrows). USP14 binding creates two parallel state-transition pathways of the proteasome. Along the substrate-inhibited pathway (red arrows), which has RPN11 blocking the substrate entrance of the OB ring before any substrate insertion takes place, USP14 trims ubiquitin chains and releases the substrate from the proteasome, thus preventing the substrate degradation (dashed turquoise arrows). Along the substrate-engaged pathway (green arrows), a substrate has already been inserted into the ATPase ring and RPN11 narrows down on the OB ring but does not block substrate translocation through the OB ring (Extended Data Fig. 7f). Although our data do not intuitively explain why USP14 trims ubiquitin until the last one on a substrate remains, the structures provide geometric constraints for polyubiquitin chain binding to both ubiquitin receptors and USP14 and suggest that ubiquitin recognition by USP14 in the proteasome requires at least one additional helper ubiquitin chain that is already anchored on a nearby ubiquitin receptor. This helper ubiquitin chain may not be available for USP14 binding but can be readily trimmed by RPN11.

However, the stabilization of substrate-inhibited states by USP14 activation does not completely exclude the possibility of substrate engagement with the ATPases. Should an unfolded or loosely folded initiation region of the substrate be available⁴⁰, substrate insertion into the AAA-ATPase channel can stochastically occur before USP14 activation by the proteasome, which presumably depends on the specific structures of substrate-ubiquitin conjugates. In the presence of ubiquitylated substrates, USP14 stimulates the ATPase rate, tightens the RP–CP interface and induces early CP gate opening^{21,22,23}, as represented in states \{\{\rm{E}\}\}_{{\rm{D4}}}\}^{{\rm{USP14}}}\}\) and \{\{\rm{E}\}\}_{{\rm{D5}}}\}^{{\rm{USP14}}}\}\). Thus, USP14 creates a second kinetic checkpoint and drives the proteasome to choose between two alternative pathways of USP14-regulated conformational transitions—one that sterically impedes substrate commitment and the other that kinetically antagonizes RPN11 by outpacing the coupling of RPN11-directed deubiquitylation and substrate translocation, rather than directly inhibiting the DUB activity of RPN11. In support of this model, all S_D-like and E_D-like states of the USP14-bound proteasome showed no ubiquitin binding to RPN11.

Previous studies have demonstrated that USP14 trims ubiquitin chains from the substrate on a millisecond time scale⁷. Consistently, all observed USP14-bound S_D-like and E_D-like states appear to present USP-bound ubiquitin in a post-cleavage state, perhaps owing to the stabilization of trimmed ubiquitin by ternary interactions of the ubiquitin–USP–OB sandwich (Fig. 2c). The lack of observation of ubiquitin-free USP14 in the proteasome suggests that the rate of ubiquitin release from USP14 on the proteasome is slow as compared with that of ubiquitin-free USP14 from the proteasome. Thus, it is conceivable that the cleaved ubiquitin chain is released only upon dissociation of the USP domain from the OB ring^{21,22,33}. This is in contrast to RPN11 in all the USP14-free E_D-like states¹⁰, which exhibit no RPN11-bound ubiquitin, implying rapid release of cleaved ubiquitin. Therefore, USP14 creates a third checkpoint, where the polyubiquitin-bound USP14 kinetically delays ubiquitin release from the proteasome and suppresses additional substrate recruitment. The third checkpoint may further control the ubiquitin-recycling function of the proteasome that is critical for regulating the free ubiquitin reservoir in cells⁴¹.

In summary, USP14 acts as an adaptive regulator of proteasomal DUB and ATPase activities that times intermediate steps of substrate processing, as USP14 bridges the gap between RPN1 and the AAA-ATPase motor in a ubiquitin-dependent, switchable fashion. USP14 interactions create three branching checkpoints on the proteasome, at the steps of initial ubiquitin recognition, substrate translocation initiation by the AAA-ATPase motor, and recycling of trimmed ubiquitin chains. This multi-checkpoint mechanism integrates catalytic and non-catalytic effects of USP14-mediated proteasome regulation into a comprehensively coordinated, elegantly timed process of substrate degradation. As partly supported by another study on the yeast Ubp6-proteasome interactions⁴², such a mechanism is expected to be conserved from yeast to human and to inform on how reversibly associated DUBs and other proteins regulate proteasome function in general (Extended Data Fig. 9i). Importantly, our time-resolved cryo-EM studies, which had been enhanced by deep learning-improved 3D classification²⁷, present an emerging paradigm that enables atomic-level visualization of hitherto inaccessible functional kinetics and complex dynamics in general.

Methods

Expression and purification of human USP14

Wild-type USP14 and its mutants were cloned into pGEX-4T vector obtained from GenScript (Nanjing, China). For purification of recombinant USP14 and mutants, BL21-CondonPlus (DE3)-RIPL cells (Shanghai Weidi) transformed with plasmids encoding wild-type or mutant USP14 were grown to an OD₆₀₀ of 0.6–0.7 in LB medium supplemented with 100 mg ml⁻¹ ampicillin. Cultures were cooled to 20 °C and induced with 0.2 mM IPTG overnight. Cells were collected by centrifugation at 3,000g for 15 min and resuspended in lysis buffer (25 mM Tris-HCl (pH 8.0), 150 mM NaCl, 0.2% NP-40, 1 mM DTT, 10% glycerol and 1× protease inhibitor cocktail). Cells were lysed by sonication and the lysate was cleared through centrifugation at 20,000g for 30 min at 4 °C. The supernatant was incubated with glutathione Sepharose 4B resin (GE Healthcare) for 3 h at 4 °C. For the purification of wild-type USP14, USP14 UBL domain (USP14-UBL) and

USP14 USP domain (USP14-USP), the resin was washed with 20 column volumes of washing buffer (25 mM Tris-HCl (pH 8.0), 150 mM NaCl, 1 mM DTT, 10% glycerol), then incubated with cleavage buffer (20 mM Tris-HCl (pH 8.0), 150 mM NaCl) containing thrombin (Sigma) overnight at 4 °C. The eluted samples were further purified on a gel-filtration column (Superdex 75, GE Healthcare) equilibrated with 25 mM Tris-HCl (pH 8.0), 150 mM NaCl, 1 mM DTT, 10% glycerol. For the purification of USP14 mutants, the resin was washed with 20 column volumes of washing buffer (25 mM Tris-HCl (pH 8.0), 300 mM NaCl, 1 mM DTT), then incubated with cleavage buffer (20 mM Tris-HCl (pH 8.0), 150 mM NaCl) containing thrombin (Sigma) overnight at 4 °C. To remove thrombin, the GST eluent was incubated with Benzamidine-Sepharose (GE Healthcare) for 30 min at 4 °C.

Expression and purification of human 26S proteasome

Hexahistidine, TEV cleavage site, biotin and hexahistidine (HTBH)-tagged human 26S proteasomes were affinity purified as described^{8,9,10} from a stable HEK 293 cell line (a gift from L. Huang). Further authentication of cell lines was not performed for this study. Mycoplasma testing was not performed for this study. In brief, HEK 293 cells were Dounce-homogenized in a lysis buffer (50 mM PBS (77.4% Na₂HPO₄, 22.6% NaH₂PO₄, pH 7.4), 5 mM MgCl₂, 5 mM ATP, 0.5% NP-40, 1 mM DTT and 10% glycerol) containing 1× protease inhibitor cocktail. The cleared lysates were incubated with Streptavidin Agarose resin (Yeasen) for 3 h at 4 °C. The resin was washed with 20 bed volumes of lysis buffer to remove endogenous USP14 and UCH37 associated with the proteasome^{7,10}. The 26S proteasomes were cleaved from the beads by TEV protease (Invitrogen) and further purified by gel filtration on a Superose 6 10/300 GL column. Western blot was used to detect RPN13 and USP14 in the proteasomes using anti-RPN13 antibody (Abcam, 1:10,000 dilution) and anti-USP14 antibody (Novus, 1:1,000 dilution). For ubiquitin–vinyl-sulfone (Ub–VS)-treated human proteasome, 1 μM Ub–VS (Boston Biochem) was added to the proteasome-binding resin and incubated for 2 h at 30 °C. Residual Ub–VS was removed by washing the beads with 30 bed volumes of wash buffer (50 mM Tris-HCl (pH7.5), 1 mM MgCl₂ and 1 mM ATP). The proteasomes were cleaved from the beads

using TEV protease (Invitrogen) and used to measure the DUB activity of USP14 using the Ub-AMC hydrolysis assay.

Preparation of polyubiquitylated Sic1^{PY}

Sic1^{PY} and WW-HECT were purified as previously described¹⁰. The PY motif (Pro-Pro-Pro-Tyr) is recognized by the WW domains of the Rsp5 family of E3 ligases. In the Sic1^{PY} construct, a PY motif was inserted to the N-terminal segment (MTPSTPPSRGTRYLA) of the Cdk inhibitor Sic1, resulting in a modified N terminus of MTPSTPPPYSRGTRYLA^{43,44} (the PY motif is underlined). Human UBE1 (plasmid obtained as a gift from C. Tang) and human UBCH5A (obtained from GenScript) were expressed as GST fusion proteins from pGEX-4T vectors. In brief, UBE1-expressing BL21-CondonPlus (DE3)-RIPL cell cultures were induced with 0.2 mM IPTG for 20 h at 16 °C, whereas UBCH5A expression was induced with 0.2 mM IPTG overnight at 18 °C. Cells were collected in lysis buffer (25 mM Tris-HCl (pH 7.5), 150 mM NaCl, 10 mM MgCl₂, 0.2% Triton-X-100, 1 mM DTT) containing 1× protease inhibitor cocktail and lysed by sonication. The cleared lysates were incubated with glutathione Sepharose 4B resin for 3 h at 4 °C and subsequently washed with 20 bed volumes of lysis buffer. The GST tag was removed by thrombin protease (Sigma) in cleavage buffer (20 mM Tris-HCl (pH 8.0), 150 mM NaCl, 1mM DTT) overnight at 4 °C. The eluted samples were further purified by gel-filtration column (Superdex 75, GE Healthcare) equilibrated with 25 mM Tris-HCl (pH 7.5), 150 mM NaCl, 1 mM DTT, 10% glycerol.

To ubiquitylate Sic1^{PY}, 1.2 μM Sic1^{PY}, 0.5 μM UBE1, 2 μM UBCH5A, 1.4 μM WW-HECT and 1 mg ml⁻¹ ubiquitin (Boston Biochem) were incubated in reaction buffer (50 mM Tris-HCl (pH 7.5), 100 mM NaCl, 10 mM MgCl₂, 2 mM ATP, 1 mM DTT and 10% glycerol) for 2 h at room temperature. His-tagged Sic1^{PY} conjugates (polyubiquitylated Sic1^{PY}, Ub_n-Sic1^{PY}) were purified by incubating with Ni-NTA resin (Qiagen) at 4 °C for 1 h. Afterwards the resin was washed with 20 column volumes of the wash buffer (50 mM Tris-HCl (pH 7.5), 100 mM NaCl, 10% glycerol). The Ub_n-Sic1^{PY} was eluted with the same buffer containing 150 mM imidazole, and finally exchanged to the storage buffer (50 mM Tris-HCl (pH 7.5), 100 mM

NaCl, 10% glycerol) using an Amicon ultrafiltration device with 30K molecular cut-off (Millipore).

Expression and purification of human RPN13

To purify human RPN13, pGEX-4T-RPN13-transformed BL21-CondonPlus (DE3)-RIPL cells were cultured to an OD₆₀₀ of 0.6 and then induced by 0.2 mM IPTG for 20 h at 16 °C. Cells were resuspended in lysis buffer (25 mM Tris-HCl (pH 7.5), 300 mM NaCl, 1 mM EDTA, 0.2% Triton-X-100, 1 mM DTT) containing 1× protease inhibitor cocktail and lysed by sonication. A 20,000g supernatant was incubated with glutathione Sepharose 4B resin (GE Healthcare) for 3 h at 4 °C. The resin was washed with 20 column volumes of washing buffer (25 mM Tris-HCl (pH 8.0), 150 mM NaCl, 1 mM DTT, 10% glycerol) and 10 column volumes of cleavage buffer (20 mM Tris-HCl (pH 8.0), 150 mM NaCl). The GST tag was cleaved by incubating with thrombin (Sigma) overnight at 4 °C. The eluted samples were further purified by gel-filtration column (Superdex 75, GE Healthcare) equilibrated with 25 mM Tris-HCl (pH 8.0), 150 mM NaCl, 1 mM DTT, 10% glycerol.

In vitro degradation assay

Purified human proteasomes (~30 nM) were incubated with RPN13 (~300 nM), Ub_n-Sic1^{PY} (~300 nM) in degradation buffer (50 mM Tris-HCl (pH 7.5), 5 mM MgCl₂ and 5 mM ATP) at 37 °C. Purified recombinant USP14 variants (~1.2 μM) were incubated with proteasome for 20 min at room temperature before initiating the degradation reaction. The reaction mixtures were incubated at 37 °C for 0, 0.5, 1.0 and 2.0 min, or 10 °C for 0, 0.5, 1.0, 5.0, 10 and 30 min, then terminated by adding SDS loading buffer and subsequently analysed by western blot using anti-T7 antibody (Abcam, 1:1,000 dilution), which was used to examine fusion protein T7-Sic1^{PY}.

Ubiquitin-AMC hydrolysis assay

Ubiquitin-AMC (Ub-AMC; Boston Biochem) hydrolysis assay was used to quantify the deubiquitylating activity of wild-type and mutant USP14 in the human proteasome. The reactions were performed in reaction buffer (50 mM

Tris-HCl (pH 7.5), 5 mM MgCl₂, 1 mM ATP, 1 mM DTT, 1 mM EDTA and 1 mg ml⁻¹ ovalbumin (Diamond)), containing 1 nM Ub–VS-treated proteasome, 0.2 µM USP14 variants and 10 nM RPN13. The reaction was initiated by adding 1 µM Ub–AMC. Ub–AMC hydrolysis was measured in a Varioskan Flash spectral scanning multimode reader (Thermo Fisher) by monitoring an increase of fluorescence excitation at 345 nm with an emission at 445 nm. For free USP14 activity, the reaction was performed using 1 µM USP14 variants and 1 µM Ub–AMC (BioVision).

ATPase activity assay

ATPase activity was quantified using malachite green phosphate assay kits (Sigma). Human proteasomes (30 nM), RPN13 (300 nM) and USP14 variants (1.2 µM) were incubated in assembly buffer (50 mM Tris-HCl (pH 7.5), 5 mM MgCl₂ and 0.5 mM ATP) for 20 min at room temperature. Ub_n–Sic1^{PY} (300 nM) was subsequently added, and the sample was incubated for 1 min at 37 °C. The reaction mixtures were mixed with malachite green buffers as described by the manufacturer (Sigma). After 30 min of room temperature incubation, the absorbance at 620 nm was determined using a Varioskan Flash spectral scanning multimode reader (Thermo Fisher).

Peptidase activity assay

Peptide hydrolysis by the human proteasomes was measured using fluorogenic substrate Suc-LLVY-AMC (MCE). Human proteasomes (1 nM) were incubated with USP14 variants (1 µM) in buffer (50 mM Tris-HCl (pH 7.5), 100 mM KCl, 0.5 mM MgCl₂, 0.1 mM ATP and 25 ng µl⁻¹ BSA) for 20 min at room temperature. 10 µM Suc-LLVY-AMC was added to the reaction mixture, which was incubated for 30 min at 37 °C in the dark. Peptide activity was measured in a Varioskan Flash spectral scanning multimode reader (Thermo Fisher) by excitation at 380 nm with an emission at 460 nm.

Microscale thermophoresis

The human proteasomes were labelled with red fluorescent dye NT-650-NHS using the Monolith NT Protein Labeling Kit (NanoTemper). After labelling, excess dye was removed by applying the sample on column B (provided in the kit) equilibrated with reaction buffer (50 mM Tris-HCl (pH 7.5), 5 mM MgCl₂ and 1 mM ATP). 0.05% Tween-20 was added to the sample before MST measurements. For interaction of NT-650-NHS-labelled proteasomes with USP14, USP14-UBL or USP14-USP, concentration series of USP14, USP14-UBL or USP14-USP were prepared using a 1:1 serial dilution of protein in reaction buffer containing 0.05% Tween-20. The range of USP14, USP14-UBL or USP14-USP concentration used began at 8 μM, with 16 serial dilution in 10-μl aliquots. The interaction was initiated by the addition of 10 μl of 30 nM NT-650-NHS-labelled proteasomes to each reaction mixture and measured by Monolith NT.115 (NanoTemper) at 20% LED excitation power and 40% MST power. To evaluate the effect of Ub_n-Sic1^{PY} on the interaction of USP14 with the proteasome, 30 nM Ub_n-Sic1^{PY} was added to the reaction mixture. Data were analysed using MO Control software provided by NanoTemper.

Cryo-EM sample preparation

To prepare cryo-EM samples, all purified proteins were exchanged to imaging buffer (50 mM Tris-HCl (pH 7.5), 5 mM MgCl₂ and 1 mM ATP). Human proteasomes (1 μM) were incubated with 10 μM RPN13, 10 μM USP14 in imaging buffer (50 mM Tris-HCl (pH 7.5), 5 mM MgCl₂ and 1 mM ATP) for 20 min at 30 °C, then cooled to 10 °C. 10 μM Ub_n-Sic1^{PY} was added to the mixture and incubated at 10 °C for 0.5, 1, 5 and 10 min. 0.005% NP-40 was added to the reaction mixture immediately before cryo-plunging. Cryo-grids made without the addition of substrate corresponded to the condition of 0 min of reaction time and were used as a baseline control for time-resolved analysis (Fig. 1e). For ATP-to-ATPγS exchange and ATPase quenching, after the reaction mixture was incubated at 10 °C for 1 min, 1 mM ATPγS was added to the reaction mixture at once, and incubated for another 1 min, then NP-40 was added to the mixture to a final concentration of 0.005% before cryo-plunging.

Cryo-EM data collection

The cryo-grids were initially screened in a 200 kV Tecnai Arctica microscope (Thermo Fisher). Good-quality grids were then transferred to a 300 kV Titan Krios G2 microscope (Thermo Fisher) equipped with the post-column BioQuantum energy filter (Gatan) connected to a K2 Summit direct electron detector (Gatan). Coma-free alignment and parallel illumination were manually optimized prior to each data collection session. Cryo-EM data were acquired automatically using SerialEM software⁴⁵ in a super-resolution counting mode with 20 eV energy slit, with the nominal defocus set in the range of -0.8 to -2.0 μm . A total exposure time of 10 s with 250 ms per frame resulted in a 40-frame movie per exposure with an accumulated dose of ~ 50 electrons per \AA^2 . The calibrated physical pixel size and the super-resolution pixel size were 1.37 \AA and 0.685 \AA , respectively. For time-resolved sample conditions, 1,781, 2,298, 15,841, 2,073 and 2,071 movies were collected for cryo-grids made with the reaction time of 0, 0.5, 1, 5, and 10 min, respectively. For the condition of exchanging ATP to ATP γ S at 1 min after substrate addition, 21,129 movies were collected.

Reference structures

Comparisons to protein structures from previous publications used the atomic models in the PDB under accession codes: 2AYN (USP domain of USP14 in its isolated form⁵), 2AYO (USP domain of USP14 bound to ubiquitin aldehyde⁵), 6MSB (state E_{A1} of substrate-engaged human proteasome¹⁰), 6MSD (state E_{A2}), 6MSE (state E_B), 6MSG (state E_{C1}), 6MSJ (state E_{D1}), 6MSK (state E_{D2}), 5VFT (state S_B of substrate-free human proteasome^{8,9}), 5VFU (state S_C), 5VFP (state S_{D1}) and 5VFR (state S_{D3}). Cryo-EM maps from previous publications used in comparison are available from EMDB under access codes EMD-9511 (USP14–UbAl-bound proteasome²⁵), EMD-3537 (Ubp6-bound proteasome map²⁶) and EMD-2995 (Ubp6–UbVS-bound proteasome²³).

Cryo-EM data processing

Drift correction and dose weighting were performed using the MotionCor2 program⁴⁶ at a super-resolution pixel size of 0.685 \AA . Drift-corrected micrographs were used for the determination of the micrograph CTF

parameters with the Gctf program⁴⁷. Particles were automatically picked on micrographs that were fourfold binned to a pixel size of 2.74 Å using an improved version of the DeepEM program⁴⁸. Micrographs screening and auto-picked particles checking were both preformed in the EMAN2 software⁴⁹. A total of 213,901, 106,564, 1,494,869, 212,685, 141,257 and 1,387,530 particles were picked for the 0 min, 0.5 min, 1 min, 5 min, 10 min and ATPγS datasets, respectively. Reference-free 2D classification and 3D classification were carried out in software packages RELION⁵⁰ version 3.1 and ROME⁵¹. Focused 3D classification, CTF and aberration refinement, and high-resolution auto-refinement were mainly done with RELION 3.1, whereas the AlphaCryo4D software²⁷ was used to analyse the conformational changes and conduct the in-depth 3D classification for time-resolved analysis. Particle subtraction and re-centering were performed using RELION 3.1 and SPIDER⁵² software. We applied a hierarchical 3D classification strategy to analyse the data (Extended Data Fig. 2), which were optimized as previously described¹⁰. The entire data-processing procedure consisted of five steps. Datasets of different conditions were processed separately at steps 1 and 2 and combined at steps 3 and 4.

Step 1: doubly capped proteasome particles were separated from singly capped ones through several rounds of 2D and 3D classification. These particles were aligned to the consensus models of the doubly and singly capped proteasome to obtain their approximate shift and angular parameters. With these parameters, each doubly capped particle was split into two pseudo-singly capped particles by re-centring the box onto the RP–CP subcomplex. Then the box sizes of pseudo-singly capped particles and true singly capped particles were both shrunk to 640 × 640 pixels with a pixel size of 0.685 Å, and down-sampled by two-fold to a pixel size of 1.37 Å for the following processing. A total of 3,429,154 particles from all datasets were obtained after this step.

Step 2: particles were aligned to the CP subcomplex through auto-refinement, followed by one round of CTF refinement to correct optical aberration (up to the fourth order), magnification anisotropy, and per-particle defocus together with per-particle astigmatism. After another run of the CP-masked auto-refinement, an alignment-skipped RP-masked 3D classification was performed to separate the S_A-like states from the S_D-like states. Poor 3D

classes showing broken 26S proteasome were removed for further analysis at this step. The RP subcomplex of the S_D-like states rotated by a large angle compared to the S_A-like states, and only in S_D-like states was the USP domain of USP14 observed to bind the OB ring of the proteasome. There were 1,774,110 particles in total in S_A-like states and 1,360,329 particles in total in S_D-like states in all datasets after this step.

Step 3: considering the particle number of some datasets were not enough to ensure high accuracy of independent 3D classification, in the following procedure we pooled particles together from all datasets except for the 0-min condition, in which the substrate was not yet added into the reaction system. For the S_D-like state, CP-masked auto-refinement was performed, followed with two rounds of CTF refinement and another run of CP-masked auto-refinement. Alignment-skipped RP-masked 3D classification was then performed, while conformational changes were analysed using AlphaCryo4D²⁷, which yielded three clusters, designated S_B-like, S_D-like, and E_D-like states. These names were correspondingly referred to their similar states in previously published studies^{9,10}. The S_A-like particles were processed in the same way, resulting in a cluster named E_A-like state; bad classes showed blurred densities in RPN10 and part of the lid. The 0-min dataset was processed independently for the lack of substrate, resulting in three classes, named S_A-like (92.8%), S_B-like (4.3%) and S_D-like (3.0%).

Step 4: particles in different clusters were individually refined with the CP masked. The CP density was then subtracted, and the particle box was recentred to the RP subcomplex and shrunk to 240 × 240 pixels, with a pixel size of 1.37 Å. For each cluster, the CP-subtracted particles were subjected to several rounds of RP-masked auto-refinement and alignment-skipped RP-masked 3D classification followed by AlphaCryo4D analysis²⁷, finally resulting in thirteen major conformational states of the USP14-bound proteasome, named \(\{\{ \rm{E} \}\} _{\{ \rm{A1} \}\}^{\{ \rm{UBL} \}}\), \(\{\{ \rm{E} \}\} _{\{ \rm{A2} \}\}.0\}^{\{ \rm{UBL} \}}\), \(\{\{ \rm{E} \}\} _{\{ \rm{A2} \}\}.1\}^{\{ \rm{UBL} \}}\), \(\{\{ \rm{S} \}\} _{\{ \rm{B} \}\}^{\{ \rm{USP14} \}}\), \(\{\{ \rm{S} \}\} _{\{ \rm{C} \}\}^{\{ \rm{USP14} \}}\), \(\{\{ \rm{S} \}\} _{\{ \rm{D4} \}\}^{\{ \rm{USP14} \}}\), \(\{\{ \rm{S} \}\} _{\{ \rm{D5} \}\}^{\{ \rm{USP14} \}}\), \(\{\{ \rm{S} \}\} _{\{ \rm{D6} \}\}^{\{ \rm{USP14} \}}\), \(\{\{ \rm{S} \}\} _{\{ \rm{D7} \}\}^{\{ \rm{USP14} \}}\), \(\{\{ \rm{S} \}\} _{\{ \rm{D8} \}\}^{\{ \rm{USP14} \}}\), \(\{\{ \rm{S} \}\} _{\{ \rm{D9} \}\}^{\{ \rm{USP14} \}}\), \(\{\{ \rm{S} \}\} _{\{ \rm{D10} \}\}^{\{ \rm{USP14} \}}\), \(\{\{ \rm{S} \}\} _{\{ \rm{D11} \}\}^{\{ \rm{USP14} \}}\), \(\{\{ \rm{S} \}\} _{\{ \rm{D12} \}\}^{\{ \rm{USP14} \}}\), \(\{\{ \rm{S} \}\} _{\{ \rm{D13} \}\}^{\{ \rm{USP14} \}}\).

$\{\{\backslash rm\{S\}\}\}_{-}\{\{\backslash rm\{D5\}\}\}^{\wedge }\{\{\backslash rm\{USP14\}\}\}\backslash), \backslash$
 $\{\{\backslash rm\{E\}\}\}_{-}\{\{\backslash rm\{D0\}\}\}^{\wedge }\{\{\backslash rm\{USP14\}\}\}\backslash), \backslash$
 $\{\{\backslash rm\{E\}\}\}_{-}\{\{\backslash rm\{D1\}\}\}^{\wedge }\{\{\backslash rm\{USP14\}\}\}\backslash), \backslash$
 $\{\{\backslash rm\{E\}\}\}_{-}\{\{\backslash rm\{D2\}\}.0\}^{\wedge }\{\{\backslash rm\{USP14\}\}\}\backslash), \backslash$
 $\{\{\backslash rm\{E\}\}\}_{-}\{\{\backslash rm\{D2\}\}.1\}^{\wedge }\{\{\backslash rm\{USP14\}\}\}\backslash), \backslash$
 $\{\{\backslash rm\{E\}\}\}_{-}\{\{\backslash rm\{D4\}\}\}^{\wedge }\{\{\backslash rm\{USP14\}\}\}\backslash)$ and \backslash
 $\{\{\backslash rm\{E\}\}\}_{-}\{\{\backslash rm\{D5\}\}\}^{\wedge }\{\{\backslash rm\{USP14\}\}\}\backslash).$ For state \backslash
 $\{\{\backslash rm\{E\}\}\}_{-}\{\{\backslash rm\{A1\}\}\}^{\wedge }\{\{\backslash rm\{UBL\}\}\}\backslash),$ particles with blurred RPN1
 were excluded for final high-resolution reconstruction. For state \backslash
 $\{\{\backslash rm\{E\}\}\}_{-}\{\{\backslash rm\{D2\}\}.1\}^{\wedge }\{\{\backslash rm\{USP14\}\}\}\backslash),$ particles with blurred
 RPN2 were excluded for final high-resolution reconstruction. These states
 exhibit remarkable conformational changes of the AAA ring and the full RP,
 as well as the interactions of the RP and USP14. Time-resolved analysis of
 conformational changes and comparison in the presence and absence of
 ATP γ S were both done after this step, by simply separating the particles for
 each state based on their time labels. Namely, the proportion of particles of
 each state at a given time point was obtained by summing up the number of
 particles for each state at the same time point and then calculating the
 fraction of particles of each state with respect to the total number of particles
 at this time point^{53,54}. Similarly, final analysis of state percentage for the
 ATP-to-ATP γ S exchange condition was done by counting the particles of
 each state under this condition, with the particles of each state used for
 separate refinement, reconstruction and comparison with those under ATP-
 only conditions (Extended Data Figs. [2c](#), [5a](#), [6i](#)).

Final refinement of each state was performed using pseudo-singly capped
 particles with the pixel size of 1.37 Å. Two types of local mask were applied
 for the auto-refinement, one focusing on the complete RP and the other
 focusing on the CP, resulting in two maps for each state, which were merged
 in Fourier space into one single map. Based on the in-plane shift and Euler
 angle of each particle from the auto-refinement, we reconstructed the two
 half-maps of each state using pseudo-singly capped particles with the pixel
 size of 0.685 Å. The Fourier shell correlation (FSC) curves of thirteen states
 were calculated from two separately refined half maps in a gold-standard
 procedure, yielding the nominal resolution ranging from 3.0 to 3.6 Å, and
 the local RP resolution ranging from 3.3 to 4.6 Å (Extended Data Figs. [2a](#),

[3b-d](#)). Before visualization, all density maps were sharpened by applying a negative B-factor ranging from -10 to -50 Å 2 .

In order to further improve the local density quality of USP14 and RPN1, which suffered from local conformational dynamics, another round of RP-masked 3D classification was done using CP-subtracted particles for some states to exclude 3D classes with blurred USP14 and RPN1. These locally improved maps were only used for visualization and adjustment of atomic models of USP14 and RPN1. For states \

\{\{\rm{E}\}\}_{{\{\rm{A1}\}}}\}^{{\{\rm{UBL}\}}}\}), \\\{\{\rm{E}\}\}_{{\{\rm{A2}\}}.0}\}^{{\{\rm{UBL}\}}}\}) and \\\{\{\rm{E}\}\}_{{\{\rm{A2}\}}.1}\}^{{\{\rm{UBL}\}}}\}), 3D classes with unblurred RPN1 and especially visible UBL on the RPN1 T2 site (including 152,802, 66,966 and 61,930 particles, respectively) were selected and refined by applying a mask on the RPN1-UBL component. The resulting RPN1-UBL density in these states were compared with previously reported E_{A1} state

(Extended Data Fig. [7e](#)). For states \

\{\{\rm{E}\}\}_{{\{\rm{D0}\}}}\}^{{\{\rm{USP14}\}}}\}) and \\\{\{\rm{E}\}\}_{{\{\rm{D2}\}}.0}\}^{{\{\rm{USP14}\}}}\}), 3D classes with unblurred RPN1 density (including 61,447 and 53,145 particles, respectively) were selected and refined to 4.1 and 4.2 Å for the RP, respectively. For state \{\{\rm{S}\}\}_{{\{\rm{C}\}}}\}^{{\{\rm{USP14}\}}}\}), \\\{\{\rm{E}\}\}_{{\{\rm{D4}\}}}\}^{{\{\rm{USP14}\}}}\}) and \\\{\{\rm{E}\}\}_{{\{\rm{D2}\}}.1}\}^{{\{\rm{USP14}\}}}\}), 3D classes with improved USP14 densities (including 34,659, 54,642, 142,814 particles, respectively) were selected and refined to 4.5, 4.2 and 3.8 Å for the RP component, respectively, for better visualization of the full-length USP14 in the proteasome.

Atomic model building and refinement

Atomic model building was based on the previously published cryo-EM structures of the human proteasome^{[10](#)}. For the CP subcomplex, initial models of the closed-gate CP and open-gate CP were respectively derived from the E_{A1} model (PDB 6MSB) and the E_{D2} model (PDB 6MSK). For the RP subcomplex, the previous E_{D2} model was used as an initial model. All subunits of the initial model were individually fitted as a rigid body into

each of the reconstructed maps with UCSF Chimera⁵⁵, followed by further adjustment of the main chain traces using Coot⁵⁶. Initial model of the full-length USP14 was first derived from a predicted one by AlphaFold⁵⁷, which was verified by comparing to a crystal structure⁵ (PDB 2AYO). The USP14 model was then merged with the initial proteasome model by independently fitting models of the USP14 UBL and USP domains as rigid bodies into the cryo-EM maps, and manually fitting the linker between the UBL and USP domains in Coot⁵⁶. Despite the presence of RPN13 in our purified proteasome (Extended Data Fig. 1g) and the addition of excessive RPN13 to saturate the proteasome, no reliable density was observed for RPN13 in all cryo-EM maps, thus precluding the atomic modelling of RPN13 and likely reflecting its highly dynamic association with the proteasome. The atomic models of subunit SEM1 and the N terminus of subunit RPN5 fitted into their corresponding local densities of lower resolution were rebuilt by considering the prediction of AlphaFold⁵⁷. The atomic model of USP14 was first rebuilt and refined against the map of state \(\{\{\backslash rm\{E\}\}\}_\{\{\backslash rm\{D2\}\}.1\}^{\{\{\backslash rm\{USP14\}\}\}}\)) with improved local USP14 density, the resulting model of which was then used as starting USP14 models to fit into the USP14 densities in other states. For some structures with partially blurred or missing subunit densities, the atomic models were revised by removing these regions, for example, the UBL domain of USP14 was removed in the models of \(\{\{\backslash rm\{E\}\}\}_\{\{\backslash rm\{D5\}\}\}^{\{\{\backslash rm\{USP14\}\}\}}\)) and \(\{\{\backslash rm\{S\}\}\}_\{\{\backslash rm\{D5\}\}\}^{\{\{\backslash rm\{USP14\}\}\}}\)). Given that the substrates were not stalled in a homogeneous location during their degradation and that substrate translocation through the proteasome is not sequence-specific, the substrate densities were modelled using polypeptide chains without assignment of amino acid sequence. For states \(\{\{\backslash rm\{E\}\}\}_\{\{\backslash rm\{A1\}\}\}^{\{\{\backslash rm\{UBL\}\}\}}\)), \(\{\{\backslash rm\{E\}\}\}_\{\{\backslash rm\{A2\}\}.0\}^{\{\{\backslash rm\{UBL\}\}\}}\)), \(\{\{\backslash rm\{E\}\}\}_\{\{\backslash rm\{A2\}\}.1\}^{\{\{\backslash rm\{UBL\}\}\}}\)), \(\{\{\backslash rm\{E\}\}\}_\{\{\backslash rm\{D0\}\}\}^{\{\{\backslash rm\{USP14\}\}\}}\)), \(\{\{\backslash rm\{E\}\}\}_\{\{\backslash rm\{D1\}\}\}^{\{\{\backslash rm\{USP14\}\}\}}\)), \(\{\{\backslash rm\{E\}\}\}_\{\{\backslash rm\{D2\}\}.0\}^{\{\{\backslash rm\{USP14\}\}\}}\)), \(\{\{\backslash rm\{E\}\}\}_\{\{\backslash rm\{D2\}\}.1\}^{\{\{\backslash rm\{USP14\}\}\}}\)) and \(\{\{\backslash rm\{E\}\}\}_\{\{\backslash rm\{D4\}\}\}^{\{\{\backslash rm\{USP14\}\}\}}\}), the nucleotide densities are of sufficient quality for differentiating ADP from ATP, which enabled us to

build the atomic models of ADP and ATP into their densities (Extended Data Fig. 6j). For other states with the local RP resolution worse than 4.0 Å, the nucleotide types or states were hypothetically inferred from their adjacent states at higher resolution with the closest structural similarity, based on the local densities, the openness of corresponding nucleotide-binding pockets as well as their homologous structural models of higher resolution if available.

After manually rebuilding, atomic models were all subjected to the real-space refinement in Phenix⁵⁸. Both stimulated annealing and global minimization were applied with non-crystallographic symmetry (NCS), rotamer and Ramachandran constraints. Partial rebuilding, model correction and density-fitting improvement in Coot⁵⁶ were then iterated after each round of atomic model refinement in Phenix⁵⁸. The refinement and rebuilding cycle were often repeated for three rounds or until the model quality reached expectation (Extended Data Table 1).

Structural analysis and visualization

All structures were analysed in Coot⁵⁶, PyMOL⁵⁹, UCSF Chimera⁵⁵, and ChimeraX⁶⁰. Inter-subunit interactions and interfacial areas were computed and analysed using the PISA server⁶¹ (https://www.ebi.ac.uk/pdbe/prot_int/pistart.html). Local resolution variations were estimated using ResMap⁶². Figures of structures were plotted in PyMOL⁵⁹, ChimeraX⁶⁰, or Coot⁵⁶. Structural alignment and comparison were performed in PyMOL⁵⁹ and ChimeraX⁶⁰.

Data reporting

No statistical methods were used to predetermine sample size. The experiments were not randomized, and investigators were not blinded to allocation during experiments and outcome assessment.

Statistical analysis

Statistical analyses of mutagenesis data were performed using two-tailed unpaired *t*-tests with SPSS v.27.0 unless otherwise indicated. Statistical

significance was assessed with a 95% confidence interval and a *P* value of < 0.05 was considered significant.

Reporting summary

Further information on research design is available in the [Nature Research Reporting Summary](#) linked to this paper.

Data availability

Cryo-EM density maps of USP14–proteasome complexes resolved in this study have been deposited in the Electron Microscopy Data Bank (EMDB) ([www.emdataresource.org](#)) under accession codes [EMD-32272](#) (\{{\rm{E}}\})_{{{\rm{UBL}}}}^{{{\rm{A1}}}}), [EMD-32273](#) (\{{\rm{E}}\})_{{{\rm{UBL}}}}^{{{\rm{A2}}}}.0), [EMD-32274](#) (\{{\rm{E}}\})_{{{\rm{UBL}}}}^{{{\rm{A2}}}}.1), [EMD-32275](#) (\{{\rm{E}}\})_{{{\rm{UBL}}}}^{{{\rm{D4}}}}), [EMD-32276](#) (\{{\rm{E}}\})_{{{\rm{UBL}}}}^{{{\rm{USP14}}}}), [EMD-32277](#) (\{{\rm{E}}\})_{{{\rm{UBL}}}}^{{{\rm{USP14}}}}), [EMD-32278](#) (\{{\rm{E}}\})_{{{\rm{UBL}}}}^{{{\rm{USP14}}}}), [EMD-32279](#) (\{{\rm{E}}\})_{{{\rm{UBL}}}}^{{{\rm{USP14}}}}), [EMD-32280](#) (\{{\rm{E}}\})_{{{\rm{UBL}}}}^{{{\rm{USP14}}}}.0), [EMD-32281](#) (\{{\rm{E}}\})_{{{\rm{UBL}}}}^{{{\rm{USP14}}}}.1), [EMD-32282](#) (\{{\rm{S}}\})_{{{\rm{USP14}}}}^{{{\rm{B}}}}), [EMD-32283](#) (\{{\rm{S}}\})_{{{\rm{USP14}}}}^{{{\rm{C}}}}), [EMD-32284](#) (\{{\rm{S}}\})_{{{\rm{USP14}}}}^{{{\rm{D4}}}}), [EMD-32285](#) (\{{\rm{S}}\})_{{{\rm{USP14}}}}^{{{\rm{D5}}}}), with the local RPN1 density improved), [EMD-32286](#) ((\{{\rm{E}}\})_{{{\rm{UBL}}}}^{{{\rm{A2}}}}.0)^{{{\rm{A1}}}}), with the local RPN1 density improved), [EMD-32287](#) ((\{{\rm{E}}\})_{{{\rm{UBL}}}}^{{{\rm{A2}}}}.1)^{{{\rm{A1}}}}), with the local RPN1 density improved), [EMD-32288](#) ((\{{\rm{E}}\})_{{{\rm{UBL}}}}^{{{\rm{D4}}}})^{{{\rm{A1}}}}), with the USP14 density improved), [EMD-32289](#) ((\{{\rm{E}}\})_{{{\rm{UBL}}}}^{{{\rm{D0}}}})^{{{\rm{A1}}}}), with the RPN1 density improved), [EMD-32290](#) ((\{{\rm{E}}\})_{{{\rm{UBL}}}}^{{{\rm{D2}}}}.0)^{{{\rm{A1}}}}), with the RPN1 density improved), [EMD-32291](#) ((\{{\rm{E}}\})_{{{\rm{UBL}}}}^{{{\rm{D2}}}}.1)^{{{\rm{A1}}}}), with the USP14 density improved) and [EMD-32292](#) ((\{{\rm{S}}\})_{{{\rm{UBL}}}}^{{{\rm{C}}}})^{{{\rm{A1}}}})

with the USP14 density improved). The corresponding coordinates have been deposited in the Protein Data Bank (PDB) (<https://www.pdb.org>) under accession codes [7W37](#) (\{\{\rm{E}\}\}_{\{\rm{A1}\}}^{\{\rm{UBL}\}}\}), [7W38](#) (\{\{\rm{E}\}\}_{\{\rm{A2}\}}.0)^{\{\rm{UBL}\}}), [7W39](#) (\{\{\rm{E}\}\}_{\{\rm{A2}\}}.1)^{\{\rm{UBL}\}}), [7W3A](#) (\{\{\rm{E}\}\}_{\{\rm{D4}\}}^{\{\rm{USP14}\}}), [7W3B](#) (\{\{\rm{E}\}\}_{\{\rm{D5}\}}^{\{\rm{USP14}\}}), [7W3C](#) (\{\{\rm{E}\}\}_{\{\rm{D0}\}}^{\{\rm{USP14}\}}), [7W3F](#) (\{\{\rm{E}\}\}_{\{\rm{D1}\}}^{\{\rm{USP14}\}}), [7W3G](#) (\{\{\rm{E}\}\}_{\{\rm{D2}\}}.0)^{\{\rm{USP14}\}}), [7W3H](#) (\{\{\rm{E}\}\}_{\{\rm{D2}\}}.1)^{\{\rm{USP14}\}}), [7W3I](#) (\{\{\rm{S}\}\}_{\{\rm{B}\}}^{\{\rm{USP14}\}}), [7W3J](#) (\{\{\rm{S}\}\}_{\{\rm{C}\}}^{\{\rm{USP14}\}}) and [7W3M](#) (\{\{\rm{S}\}\}_{\{\rm{D4}\}}^{\{\rm{USP14}\}}) and [7W3M](#) (\{\{\rm{S}\}\}_{\{\rm{D5}\}}^{\{\rm{USP14}\}}). Uncropped versions of all gels and blots are provided in Supplementary Fig. 1. All other data are available from the corresponding author upon reasonable request. [Source data](#) are provided with this paper.

References

1. Mao, Y. In *Macromolecular Protein Complexes III: Structure and Function* (eds Harris J. R. & Marles-Wright, J.) Ch. 1 (Springer, 2021).
2. de Poot, S. A. H., Tian, G. & Finley, D. Meddling with fate: the proteasomal deubiquitinating enzymes. *J. Mol. Biol.* **429**, 3525–3545 (2017).
3. Leggett, D. S. et al. Multiple associated proteins regulate proteasome structure and function. *Mol. Cell* **10**, 495–507 (2002).
4. Borodovsky, A. et al. A novel active site-directed probe specific for deubiquitylating enzymes reveals proteasome association of USP14. *EMBO J.* **20**, 5187–5196 (2001).
5. Hu, M. et al. Structure and mechanisms of the proteasome-associated deubiquitinating enzyme USP14. *EMBO J.* **24**, 3747–3756 (2005).

6. Lee, B. H. et al. Enhancement of proteasome activity by a small-molecule inhibitor of USP14. *Nature* **467**, 179–184 (2010).
7. Lee, B. H. et al. USP14 deubiquitinates proteasome-bound substrates that are ubiquitinated at multiple sites. *Nature* **532**, 398–401 (2016).
8. Chen, S. et al. Structural basis for dynamic regulation of the human 26S proteasome. *Proc. Natl Acad. Sci. USA* **113**, 12991–12996 (2016).
9. Zhu, Y. et al. Structural mechanism for nucleotide-driven remodeling of the AAA-ATPase unfoldase in the activated human 26S proteasome. *Nat. Commun.* **9**, 1360 (2018).
10. Dong, Y. et al. Cryo-EM structures and dynamics of substrate-engaged human 26S proteasome. *Nature* **565**, 49–55 (2019).
11. Yao, T. & Cohen, R. E. A cryptic protease couples deubiquitination and degradation by the proteasome. *Nature* **419**, 403–407 (2002).
12. Verma, R. et al. Role of Rpn11 metalloprotease in deubiquitination and degradation by the 26S proteasome. *Science* **298**, 611–615 (2002).
13. Worden, E. J., Dong, K. C. & Martin, A. An AAA motor-driven mechanical switch in Rpn11 controls deubiquitination at the 26S proteasome. *Mol. Cell* **67**, 799–811.e798 (2017).
14. Wang, D., Ma, H., Zhao, Y. & Zhao, J. Ubiquitin-specific protease 14 is a new therapeutic target for the treatment of diseases. *J. Cell. Physiol.* **236**, 3396–3405 (2021).
15. Harrigan, J. A., Jacq, X., Martin, N. M. & Jackson, S. P. Deubiquitylating enzymes and drug discovery: emerging opportunities. *Nat. Rev. Drug Discov.* **17**, 57–78 (2018).
16. Wertz, I. E. & Wang, X. From discovery to bedside: targeting the ubiquitin system. *Cell Chem. Biol.* **26**, 156–177 (2019).
17. Wang, X. et al. The proteasome deubiquitinase inhibitor VLX1570 shows selectivity for ubiquitin-specific protease-14 and induces

apoptosis of multiple myeloma cells. *Sci. Rep.* **6**, 26979 (2016).

18. Hanna, J. et al. Deubiquitinating enzyme Ubp6 functions noncatalytically to delay proteasomal degradation. *Cell* **127**, 99–111 (2006).
19. Crosas, B. et al. Ubiquitin chains are remodeled at the proteasome by opposing ubiquitin ligase and deubiquitinating activities. *Cell* **127**, 1401–1413 (2006).
20. Kim, H. T. & Goldberg, A. L. The deubiquitinating enzyme Usp14 allosterically inhibits multiple proteasomal activities and ubiquitin-independent proteolysis. *J. Biol. Chem.* **292**, 9830–9839 (2017).
21. Peth, A., Kukushkin, N., Bosse, M. & Goldberg, A. L. Ubiquitinated proteins activate the proteasomal ATPases by binding to Usp14 or Uch37 homologs. *J. Biol. Chem.* **288**, 7781–7790 (2013).
22. Peth, A., Besche, H. C. & Goldberg, A. L. Ubiquitinated proteins activate the proteasome by binding to Usp14/Ubp6, which causes 20S gate opening. *Mol. Cell* **36**, 794–804 (2009).
23. Bashore, C. et al. Ubp6 deubiquitinase controls conformational dynamics and substrate degradation of the 26S proteasome. *Nat. Struct. Mol. Biol.* **22**, 712–719 (2015).
24. Aufderheide, A. et al. Structural characterization of the interaction of Ubp6 with the 26S proteasome. *Proc. Natl Acad. Sci. USA* **112**, 8626–8631 (2015).
25. Huang, X., Luan, B., Wu, J. & Shi, Y. An atomic structure of the human 26S proteasome. *Nat. Struct. Mol. Biol.* **23**, 778–785 (2016).
26. Wehmer, M. et al. Structural insights into the functional cycle of the ATPase module of the 26S proteasome. *Proc. Natl Acad. Sci. USA* **114**, 1305–1310 (2017).
27. Wu, Z. et al. Hidden dynamics of human proteasome autoregulation discovered by cryo-EM data-driven deep learning. Preprint at

<https://doi.org/10.1101/2020.12.22.423932> (2021).

28. Schweitzer, A. et al. Structure of the human 26S proteasome at a resolution of 3.9 Å. *Proc. Natl Acad. Sci. USA* **113**, 7816–7821 (2016).
29. Haselbach, D. et al. Long-range allosteric regulation of the human 26S proteasome by 20S proteasome-targeting cancer drugs. *Nat. Commun.* **8**, 15578 (2017).
30. de la Pena, A. H., Goodall, E. A., Gates, S. N., Lander, G. C. & Martin, A. Substrate-engaged 26S proteasome structures reveal mechanisms for ATP-hydrolysis-driven translocation. *Science* **362**, eaav0725 (2018).
31. Eisele, M. R. et al. Expanded coverage of the 26S proteasome conformational landscape reveals mechanisms of peptidase gating. *Cell Rep.* **24**, 1301–1315.e1305 (2018).
32. Shi, Y. et al. Rpn1 provides adjacent receptor sites for substrate binding and deubiquitination by the proteasome. *Science* **351**, eaad9421 (2016).
33. Kuo, C. L. & Goldberg, A. L. Ubiquitinated proteins promote the association of proteasomes with the deubiquitinating enzyme Usp14 and the ubiquitin ligase Ube3c. *Proc. Natl Acad. Sci. USA* **114**, E3404–E3413 (2017).
34. Kim, H. T. & Goldberg, A. L. UBL domain of Usp14 and other proteins stimulates proteasome activities and protein degradation in cells. *Proc. Natl Acad. Sci. USA* **115**, E11642–E11650 (2018).
35. Collins, G. A. & Goldberg, A. L. Proteins containing ubiquitin-like (Ubl) domains not only bind to 26S proteasomes but also induce their activation. *Proc. Natl Acad. Sci. USA* **117**, 4664–4674 (2020).
36. Zhang, S. & Mao, Y. AAA+ ATPases in protein degradation: structures, functions and mechanisms. *Biomolecules* **10**, 629 (2020).
37. Saha, A. & Warshel, A. Simulating the directional translocation of a substrate by the AAA+ motor in the 26S proteasome. *Proc. Natl Acad. Sci. USA* **118**, e2104245118 (2021).

38. Erales, J., Hoyt, M. A., Troll, F. & Coffino, P. Functional asymmetries of proteasome translocase pore. *J. Biol. Chem.* **287**, 18535–18543 (2012).
39. Beckwith, R., Estrin, E., Worden, E. J. & Martin, A. Reconstitution of the 26S proteasome reveals functional asymmetries in its AAA+ unfoldase. *Nat. Struct. Mol. Biol.* **20**, 1164–1172 (2013).
40. Prakash, S., Tian, L., Ratliff, K. S., Lehotzky, R. E. & Matouschek, A. An unstructured initiation site is required for efficient proteasome-mediated degradation. *Nat. Struct. Mol. Biol.* **11**, 830–837 (2004).
41. Chen, P. C. et al. The proteasome-associated deubiquitinating enzyme Usp14 is essential for the maintenance of synaptic ubiquitin levels and the development of neuromuscular junctions. *J. Neurosci.* **29**, 10909–10919 (2009).
42. Hung, K. Y. S. et al. Allosteric control of Ubp6 and the proteasome via a bidirectional switch. *Nat. Commun.* **13**, 838 (2022).
43. Verma, R., McDonald, H., Yates, J. R. 3rd & Deshaies, R. J. Selective degradation of ubiquitinated Sic1 by purified 26S proteasome yields active S phase cyclin–Cdk. *Mol. Cell* **8**, 439–448 (2001).
44. Verma, R., Oania, R., Graumann, J. & Deshaies, R. J. Multiubiquitin chain receptors define a layer of substrate selectivity in the ubiquitin–proteasome system. *Cell* **118**, 99–110 (2004).
45. Mastronarde, D. N. Automated electron microscope tomography using robust prediction of specimen movements. *J. Struct. Biol.* **152**, 36–51 (2005).
46. Zheng, S. Q. et al. MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat. Methods* **14**, 331–332 (2017).
47. Zhang, K. Gctf: real-time CTF determination and correction. *J. Struct. Biol.* **193**, 1–12 (2016).

48. Zhu, Y., Ouyang, Q. & Mao, Y. A deep convolutional neural network approach to single-particle recognition in cryo-electron microscopy. *BMC Bioinformatics* **18**, 348 (2017).
49. Tang, G. et al. EMAN2: an extensible image processing suite for electron microscopy. *J. Struct. Biol.* **157**, 38–46 (2007).
50. Zivanov, J. et al. New tools for automated high-resolution cryo-EM structure determination in RELION-3. *eLife* **7**, e42166 (2018).
51. Wu, J. et al. Massively parallel unsupervised single-particle cryo-EM data clustering via statistical manifold learning. *PLoS ONE* **12**, e0182130 (2017).
52. Shaikh, T. R. et al. SPIDER image processing for single-particle reconstruction of biological macromolecules from electron micrographs. *Nat. Protoc.* **3**, 1941–1974 (2008).
53. Fischer, N., Konevega, A. L., Wintermeyer, W., Rodnina, M. V. & Stark, H. Ribosome dynamics and tRNA movement by time-resolved electron cryomicroscopy. *Nature* **466**, 329–333 (2010).
54. Kaledhonkar, S. et al. Late steps in bacterial translation initiation visualized using time-resolved cryo-EM. *Nature* **570**, 400–404 (2019).
55. Pettersen, E. F. et al. UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).
56. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D* **60**, 2126–2132 (2004).
57. Tunyasuvunakool, K. et al. Highly accurate protein structure prediction for the human proteome. *Nature* **596**, 590–596 (2021).
58. Adams, P. D. et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D* **66**, 213–221 (2010).

59. The PyMOL Molecular Graphics System Version 2.0 (Schrödinger, LLC).
60. Goddard, T. D. et al. UCSF ChimeraX: Meeting modern challenges in visualization and analysis. *Protein Sci.* **27**, 14–25 (2018).
61. Krissinel, E. & Henrick, K. Inference of macromolecular assemblies from crystalline state. *J Mol. Biol.* **372**, 774–797 (2007).
62. Kucukelbir, A., Sigworth, F. J. & Tagare, H. D. Quantifying the local resolution of cryo-EM density maps. *Nat. Methods* **11**, 63–65 (2014).

Acknowledgements

We thank Y. Saeki for the plasmids expressing Sic1^{PY} and WWHECT; L. Huang for the proteasome-expressing cell lines; C. Tang for the plasmid expressing UBE1; X. Li, Z. Guo, X. Pei, B. Shao, G. Wang, Y. Ma, C. Fan, B. Liu and L. Lai for technical assistance; and A. Goldberg and Q. Ouyang for discussions and comments. This work was funded in part by grants from the Beijing Natural Science Foundation (Z180016/Z18J008) and the National Natural Science Foundation of China (11774012, 12125401) to Y.M., and by a grant from the National Institute of Health (R01 GM043601) to D.F. The cryo-EM data were collected at the Cryo-EM Core Facility Platform and Laboratory of Electron Microscopy at Peking University. The data processing was performed in the High-Performance Computing Platform, the Weiming No. 1 and Life Science No. 1 supercomputing systems, at Peking University.

Author information

Author notes

1. These authors contributed equally: Shuwen Zhang, Shitao Zou

Affiliations

1. State Key Laboratory for Artificial Microstructures and Mesoscopic Physics, School of Physics, Peking University, Beijing, China

Shuwen Zhang, Shitao Zou, Deyao Yin, Lihong Zhao, Zhaolong Wu & Youdong Mao

2. Peking-Tsinghua Joint Center for Life Sciences, Peking University, Beijing, China

Shuwen Zhang, Shitao Zou, Deyao Yin, Lihong Zhao, Zhaolong Wu & Youdong Mao

3. Department of Cell Biology, Harvard Medical School, Boston, MA, USA

Daniel Finley

4. Center for Quantitative Biology, Peking University, Beijing, China

Zhaolong Wu & Youdong Mao

5. National Biomedical Imaging Center, Peking University, Beijing, China

Youdong Mao

Contributions

Y.M. conceived this study. S. Zou purified the protein complexes and prepared the cryo-EM samples. S. Zou and L.Z. conducted the biochemical experiments and mutagenesis study. S. Zhang, S. Zou, and D.Y. collected cryo-EM data and analysed the experimental cryo-EM datasets. S. Zhang refined the density maps. S. Zhang and Y.M. built and refined the atomic models. D.F. and Z.W. contributed to the analysis of the data. Y.M. supervised this study, analysed the data and wrote the manuscript with input from all authors.

Corresponding author

Correspondence to [Youdong Mao](#).

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 Protein purification and cryo-EM imaging.

a, The human 26S proteasome was purified through gel-filtration column (Superose 6 10/300 GL). **b**, Native gel analysis of the human 26S proteasome from (a). **c**, FPLC purification of human USP14 on Superdex 75 10/300 GL column. **d–f**, SDS-PAGE and Coomassie blue stain analysis of purified USP14 (**d**), Sic1^{PY}, UBE1, UBCH5A, WW-HECT, human RPN13 (**e**), USP14 UBL and USP domains (**f**). **g** and **h**, Western blot was used to evaluate the content of RPN13 (**g**) and USP14 (**h**) in the purified human proteasomes. The results indicate the presence of RPN13 and the absence of USP14 in the purified human 26S proteasome. **i**, Western blot was used to verify polyubiquitylation of Sic1^{PY} ($\text{Ub}_n\text{-Sic1}^{\text{PY}}$) using anti-T7 antibody, indicating that most Sic1^{PY} was ubiquitylated. **j**, *In vitro* degradation of $\text{Ub}_n\text{-Sic1}^{\text{PY}}$ by the purified 26S proteasome at 10 °C in the absence and presence of USP14. The concentration and ratio of each component was same as cryo-EM sample preparation. The experiments were repeated three times. Samples in (i) and (j) were analyzed by SDS-PAGE/Western blot using anti-T7 antibody. W/O, the proteasome without binding to USP14. WT, the wildtype USP14-bound proteasome. **k**, Kinetics of $\text{Ub}_n\text{-Sic1}^{\text{PY}}$ degradation was plotted by measuring $\text{Ub}_n\text{-Sic1}^{\text{PY}}$ density in (j) using ImageJ software. Each point is representative of three independent experiments. Data are presented as mean ± s.d. **l–o**, MST analysis of USP14 binding to the human proteasome. A dissociation constant of 94.6 ± 27.1 nM (**l**, full-length USP14 in the absence of $\text{Ub}_n\text{-Sic1}^{\text{PY}}$), 137 ± 33.4 nM (**m**, USP14 UBL domain only), 135 ± 22.9 nM (**n**, USP14 USP domain only) and 43.9 ± 24.7 nM (**o**, full-length USP14 in the presence of $\text{Ub}_n\text{-Sic1}^{\text{PY}}$) were calculated from three independent experiments (shown as mean ± s.d.). **p** and **q**, Typical motion-corrected cryo-EM micrographs (left) of the substrate-engaged human USP14–proteasome complex in the presence of 1 mM ATP (**p**) or after ATP-to-ATPγS exchange (**q**). Power spectrum evaluation of the corresponding micrographs are shown on the right. The exact numbers of micrographs collected under different experimental conditions are provided in Extended Data Fig. 2a. Each experiment was repeated independently at least five times with similar results. **r**,

Comparison of percent population of each conformational state in the presence of 1 mM ATP or ATP-to-ATP γ S exchange in 1 min after mixing the substrate with the USP14-bound proteasome. Our cryo-EM analysis suggests that ATP-to-ATP γ S exchange enriches states \(\{\{\backslash\text{rm}\{ E \}\}\}_{-}\{\{\backslash\text{rm}\{ A2 \}\}.1\}^{\{\{\backslash\text{rm}\{ UBL \}\}\}}\), \(\{\{\backslash\text{rm}\{ S \}\}\}_{-}\{\{\backslash\text{rm}\{ C \}\}\}^{\{\{\backslash\text{rm}\{ USP14 \}\}\}}\), \(\{\{\backslash\text{rm}\{ S \}\}\}_{-}\{\{\backslash\text{rm}\{ D5 \}\}\}^{\{\{\backslash\text{rm}\{ USP14 \}\}\}}\), \(\{\{\backslash\text{rm}\{ E \}\}\}_{-}\{\{\backslash\text{rm}\{ D1 \}\}\}^{\{\{\backslash\text{rm}\{ USP14 \}\}\}}\), \(\{\{\backslash\text{rm}\{ E \}\}\}_{-}\{\{\backslash\text{rm}\{ D2 \}\}.1\}^{\{\{\backslash\text{rm}\{ USP14 \}\}\}}\), and \(\{\{\backslash\text{rm}\{ E \}\}\}_{-}\{\{\backslash\text{rm}\{ D4 \}\}\}^{\{\{\backslash\text{rm}\{ USP14 \}\}\}}\), but reduces states \(\{\{\backslash\text{rm}\{ E \}\}\}_{-}\{\{\backslash\text{rm}\{ A1 \}\}\}^{\{\{\backslash\text{rm}\{ UBL \}\}\}}\), \(\{\{\backslash\text{rm}\{ E \}\}\}_{-}\{\{\backslash\text{rm}\{ A2 \}\}.0\}^{\{\{\backslash\text{rm}\{ UBL \}\}\}}\), \(\{\{\backslash\text{rm}\{ S \}\}\}_{-}\{\{\backslash\text{rm}\{ B \}\}\}^{\{\{\backslash\text{rm}\{ USP14 \}\}\}}\), and \(\{\{\backslash\text{rm}\{ E \}\}\}_{-}\{\{\backslash\text{rm}\{ D2 \}\}.0\}^{\{\{\backslash\text{rm}\{ USP14 \}\}\}}\). The particle numbers used to derive this plot are provided in Extended Data Fig. 2c. For gel source data, see Supplementary Fig. 1

Source data

Extended Data Fig. 2 Cryo-EM data processing workflow and time-resolved analysis.

a, The workflow diagram illustrates the major steps of our focused 3D classification strategy for cryo-EM data analysis. Particle numbers after 3D classification and final reconstruction and the resolutions of the complete RP-CP and RP-masked reconstructions of each state are labelled. **b**, 3D classification of the dataset corresponding to 0 minute before substrate addition as an overall control. **c**, Time-resolved analysis of all states by restoring the time label (for the buffer condition with 1 mM ATP) or ATP γ S label (for the condition with ATP-to-ATP γ S exchange). The percentages were computed using the total particle number corresponding to a given time point.

Extended Data Fig. 3 Cryo-EM reconstructions and resolution measurement.

a, Local resolution estimation of the RP reconstructions of thirteen states calculated by ResMap⁶². Each state is shown in two orthogonal orientations, with the second orientation (top view) shown in two different cross-sections at the AAA (middle row) and OB (lower row) domains. All local resolutions are shown with the same color bar in the upper right inset. **b** and **c**, Gold-standard Fourier shell correlation (FSC) plots of the complete RP-CP maps of all states calculated without (**b**) or with (**c**) masking the separately refined half-maps. **d**, Gold-standard FSC plots of the RP-masked reconstructions of all states. The RP maps were refined by focusing the mask on the RP subcomplex. **e**, Model-map FSC plots calculated by Phenix⁵⁸ between each RP-CP map (masked) and its corresponding atomic model. For each state, separately refined RP and CP maps (using RP and CP masks, respectively) were merged in Fourier space into a single RP-CP map, which was used for the model-map FSC calculation. The same color code is used in (**b–e**).

Extended Data Fig. 4 Cryo-EM maps and quality assessment.

a, Gallery of all refined cryo-EM maps not shown in the main figures. **b–f**, Typical high-resolution cryo-EM densities (mesh) of secondary structures superimposed with their atomic models. Different subunits of the proteasome are shown in (**b**), where the substrates shown in the middle panel are modelled using polypeptide chains without assignment of amino acid sequence. **c**, High-resolution cryo-EM densities of the two domains of USP14 in state $\backslash(\{ \{ \backslash\text{rm}\{ E \} \} \} _\{ \{ \backslash\text{rm}\{ D2 \} \} .1 \} ^\{ \{ \backslash\text{rm}\{ USP14 \} \} \})$. **d**, High-resolution cryo-EM densities of USP14 BL1 and BL2 loops and USP-OB interface in state $\backslash(\{ \{ \backslash\text{rm}\{ E \} \} \} _\{ \{ \backslash\text{rm}\{ D2 \} \} .1 \} ^\{ \{ \backslash\text{rm}\{ USP14 \} \} \})$. **e**, High-resolution cryo-EM densities of the BL3 loop, ubiquitin and representative secondary structure elements in the USP domain of USP14 in state $\backslash(\{ \{ \backslash\text{rm}\{ E \} \} \} _\{ \{ \backslash\text{rm}\{ D4 \} \} \} ^\{ \{ \backslash\text{rm}\{ USP14 \} \} \})$. **f**, High-resolution cryo-EM densities of USP-AAA interfaces in state $\backslash(\{ \{ \backslash\text{rm}\{ E \} \} \} _\{ \{ \backslash\text{rm}\{ D4 \} \} \} ^\{ \{ \backslash\text{rm}\{ USP14 \} \} \})$.

Extended Data Fig. 5 Structural comparison of the proteasome under distinct conditions.

a, Comparison of cryo-EM reconstructions of the thirteen states of USP14-proteasome complex from the conditions of including 1 mM ATP (ATP-only, no ATP γ S in the degradation buffer) (orange) and with ATP-to-ATP γ S exchange at 1 min after substrate addition (blue). Particle number and the RP resolution for each class are labelled below each panel of structural comparison. The results show that consistent features of the same states were obtained from the two conditions, although the N-terminal densities of RPN5 are slightly stronger under the condition with ATP-to-ATP γ S exchange. **b**, Comparison of the three major states at 0 min (before substrate addition and after USP14 was bound to the proteasome) with those after the substrate was mixed with the USP14-proteasome complex. For visual clarity, all maps were low-pass filtered to 8 Å. **c**, Comparison of state \(\{\{\backslash\text{rm}\{S\}\}\}_{\{\{\backslash\text{rm}\{B\}\}\}}^{\{\{\backslash\text{rm}\{USP14\}\}\}}\backslash\) with the USP14-UbAl-bound proteasome map EMD-9511 (panel 1)²⁵ and of state \(\{\{\backslash\text{rm}\{E\}\}\}_{\{\{\backslash\text{rm}\{D2\}\}.1\}}^{\{\{\backslash\text{rm}\{USP14\}\}\}}\backslash\) with the Ubp6-bound proteasome map EMD-3537 (panel 2)²⁶ and Ubp6-UbVS-bound proteasome map EMD-2995 (panel 3)²³. Due to the low-resolution nature of the previously published maps, the cryo-EM maps of states \(\{\{\backslash\text{rm}\{S\}\}\}_{\{\{\backslash\text{rm}\{B\}\}\}}^{\{\{\backslash\text{rm}\{USP14\}\}\}}\backslash\) and \(\{\{\backslash\text{rm}\{E\}\}\}_{\{\{\backslash\text{rm}\{D2\}\}.1\}}^{\{\{\backslash\text{rm}\{USP14\}\}\}}\backslash\) were low-pass filtered to 8 Å for the visual clarity of comparison. **d**, Structural comparison of state \(\{\{\backslash\text{rm}\{E\}\}\}_{\{\{\backslash\text{rm}\{D4\}\}\}}^{\{\{\backslash\text{rm}\{USP14\}\}\}}\backslash\) with state E_B (6MSE) suggests that they have comparable ATPase conformations with defined differences in RPT6 (shown in panel **e**), whereas the lid subcomplexes are in very different conformations with large rotations. **e**, Structural comparison of RPT6 in states \(\{\{\backslash\text{rm}\{E\}\}\}_{\{\{\backslash\text{rm}\{D4\}\}\}}^{\{\{\backslash\text{rm}\{USP14\}\}\}}\backslash\) and E_B shows that the RPT6 pore-1 loop, highlighted by transparent sphere representation of Phe223, is moved about 7.2 Å toward the substrate in state \(\{\{\backslash\text{rm}\{E\}\}\}_{\{\{\backslash\text{rm}\{D4\}\}\}}^{\{\{\backslash\text{rm}\{USP14\}\}\}}\backslash\) relative to state E_B. The right panel show the two RPT6 structures superimposed after aligning the entire ATPase motor subcomplex structures together (as shown in the right of panel **d**). **f**, Comparison of the RP and ATPase structures in different states and previously published cryo-EM structures^{8,9,10}. These structures are aligned together against their CPs. Each pair of compared structures are shown in two orthogonal orientations, with the root-mean-squared-

deviation (RMSD) values for the ATPase components shown below each panel of structural comparison. Previous USP14-free structures (PDB ID) used for the comparison include substrate-free states S_B (5VFT), S_C (5VFU), S_{D1} (5VFP), S_{D3} (5VFR) and substrate-bound states E_{A1} (6MSB), E_{A2} (6MSD), E_B (6MSE), E_{C1} (6MSG), E_{D1} (6MSJ), and E_{D2} (6MSK).

Extended Data Fig. 6 Substrate interactions with the pore loops in the AAA-ATPase motor, the RP-CP interfaces and the nucleotide densities in different USP14-bound states.

a–f, Cryo-EM densities of the ATPase motor bound to the substrate. Substrate densities are colored in red. Right insets show the zoomed-in side views of the substrate interacting with the pore loops of the ATPases. The substrate in each state is modelled as a polypeptide backbone structure and is represented with red sticks. The ATPases and substrates are rendered as surface (left) and mesh (right) representations. **g**, Top views of the ATPase motor structures of all states not shown in the main figure are shown in cartoon representations. Nucleotides are shown in sphere representations. The sphere representations of ADP and ATP are in blue and in red, respectively. The structures are aligned together against their CP components. Top right, color codes of subunits used in all panels. **h**, Comparison of the RP–CP interface and RPT C-terminal tail insertions into the α -pockets of the CP in different states. The cryo-EM densities of the CP subcomplexes are shown as grey surface representations, while the RPT C-tails are colored in teal blue. **i**, Comparison of the nucleotide densities in the cryo-EM reconstructions of USP14-bound proteasome under two different nucleotide conditions. The side-by-side comparison of the nucleotide densities in two best resolved states show that the ATP-to-ATP γ S exchange did not change the observed nucleotide states under our experimental conditions relative to that only used 1 mM ATP in the degradation buffer. **j**, Comparison of the nucleotide densities in thirteen distinct states of UBP14-bound proteasome. The nucleotide densities fitted with atomic models are shown in blue mesh representation. All close-up views were directly screen-copied from Coot after atomic modelling into the density maps without modification. At the contour level commonly used for atomic modelling, the potential nucleotide densities in the apo-like subunits mostly disappear,

although they can occasionally appear as partial nucleotide shapes at a much lower contour level. The states with limited local resolutions are hypothetically assigned for nucleotide types based on the densities, the openness of corresponding nucleotide-binding pockets as well as their homologous structural models of higher resolution from states with similar conformations.

Extended Data Fig. 7 Structural comparison of the USP14-bound proteasome of different states.

a, Structural comparison of states \

\(\{\{\backslash\text{rm}\{\text{E}\}\}\}_\{\{\backslash\text{rm}\{\text{A1}\}\}\}^\{\{\backslash\text{rm}\{\text{UBL}\}\}\}\), \(\{\{\backslash\text{rm}\{\text{E}\}\}\}_\{\{\backslash\text{rm}\{\text{A2}\}\}\}.0\}^\{\{\backslash\text{rm}\{\text{UBL}\}\}\}\) and \(\{\{\backslash\text{rm}\{\text{E}\}\}\}_\{\{\backslash\text{rm}\{\text{A2}\}\}\}.1\}^\{\{\backslash\text{rm}\{\text{UBL}\}\}\}\) shows the ubiquitin transfer from the RPT4-RPT5 coiled-coil (CC) domain to RPN11. The cryo-EM densities rendered as grey mesh representations are low-pass-filtered to 8 Å for visual clarity of comparison. **b**, Structural comparison of ubiquitin-RPN11-RPT5 interaction in state \

\(\{\{\backslash\text{rm}\{\text{E}\}\}\}_\{\{\backslash\text{rm}\{\text{A2}\}\}\}.0\}^\{\{\backslash\text{rm}\{\text{UBL}\}\}\}\) and \(\{\{\backslash\text{rm}\{\text{E}\}\}\}_\{\{\backslash\text{rm}\{\text{A2}\}\}\}.1\}^\{\{\backslash\text{rm}\{\text{UBL}\}\}\}\). Cryo-EM densities rendered transparent surfaces are superimposed with the atomic models.

The overall conformation of \

\(\{\{\backslash\text{rm}\{\text{E}\}\}\}_\{\{\backslash\text{rm}\{\text{A2}\}\}\}.0\}^\{\{\backslash\text{rm}\{\text{UBL}\}\}\}\) resides between \(\{\{\backslash\text{rm}\{\text{E}\}\}\}_\{\{\backslash\text{rm}\{\text{A1}\}\}\}^\{\{\backslash\text{rm}\{\text{UBL}\}\}\}\) and \(\{\{\backslash\text{rm}\{\text{E}\}\}\}_\{\{\backslash\text{rm}\{\text{A2}\}\}\}.1\}^\{\{\backslash\text{rm}\{\text{UBL}\}\}\}\). Both states \(\{\{\backslash\text{rm}\{\text{E}\}\}\}_\{\{\backslash\text{rm}\{\text{A2}\}\}\}.0\}^\{\{\backslash\text{rm}\{\text{UBL}\}\}\}\) and \(\{\{\backslash\text{rm}\{\text{E}\}\}\}_\{\{\backslash\text{rm}\{\text{A2}\}\}\}.1\}^\{\{\backslash\text{rm}\{\text{UBL}\}\}\}\) exhibited RPN11-bound ubiquitin and no substrate densities in the AAA-ATPase motor. A short stretch of ubiquitin-linked substrate density is bound to the cleft between RPN11 and the OB ring in \(\{\{\backslash\text{rm}\{\text{E}\}\}\}_\{\{\backslash\text{rm}\{\text{A2}\}\}\}.1\}^\{\{\backslash\text{rm}\{\text{UBL}\}\}\}\) but not \(\{\{\backslash\text{rm}\{\text{E}\}\}\}_\{\{\backslash\text{rm}\{\text{A2}\}\}\}.0\}^\{\{\backslash\text{rm}\{\text{UBL}\}\}\}\). In both states, the RPT5 N-loop (residues 99-119) pairs with one side of the Insert-1 β-hairpin of RPN11, the other side of which is paired with the C terminus of ubiquitin. They together form a four-stranded β-sheet, a feature that was previously visualized at atomic level only in state E_B of the USP14-free proteasome¹⁰. **c**, Structural comparison of the RP and ATPase between \(\{\{\backslash\text{rm}\{\text{E}\}\}\}_\{\{\backslash\text{rm}\{\text{A2}\}\}\}.0\}^\{\{\backslash\text{rm}\{\text{UBL}\}\}\}\) and \

(\{{\rm{E}}\}\}_{\{{\rm{A2}}\}}.1\}^{\{{\rm{UBL}}\}}\}) and between \(\{{\rm{E}}\}\}_{\{{\rm{A2}}\}}.1\}^{\{{\rm{UBL}}\}}\}) and E_B (PDB ID 6MSE).

d, Structural comparison of the RP and ATPase between \(\{{\rm{E}}\}\}_{\{{\rm{D4}}\}}^{\{{\rm{USP14}}\}}\}) and \(\{{\rm{S}}\}\}_{\{{\rm{D4}}\}}^{\{{\rm{USP14}}\}}\}) and between \(\{{\rm{E}}\}\}_{\{{\rm{D5}}\}}^{\{{\rm{USP14}}\}}\}) and \(\{{\rm{S}}\}\}_{\{{\rm{D5}}\}}^{\{{\rm{USP14}}\}}\}). The root-mean-squared-deviation (RMSD) values for the ATPase components are shown below each panel of structural comparison in **c** and **d**. **e**, Binding of UBL of USP14 to the T2 site of RPN1 in state \(\{{\rm{E}}\}\}_{\{{\rm{A1}}\}}^{\{{\rm{UBL}}\}}\}) as compared to the binding of ubiquitin to RPN1 in USP14-free state E_{A1} from previous studies¹⁰.

Right inset shows the cryo-EM density of UBL-bound RPN1 in \(\{{\rm{E}}\}\}_{\{{\rm{A1}}\}}^{\{{\rm{UBL}}\}}\}). **f**, Comparison of the RPN11-OB ring interface in different states. Left panel, the interface in the E_A-like states shows an open OB ring for substrate entrance. Middle panel, the substrate entrance of the OB ring is blocked by RPN11 in the substrate-inhibited states. Right panel, RPN11 is rotated outward slightly by ~5 Å to make way for substrate translocation through the OB ring in state \(\{{\rm{E}}\}\}_{\{{\rm{D2}}\}}.1\}^{\{{\rm{USP14}}\}}\}) (grey cartoon) as compared to those in the substrate-inhibited states (colorful cartoon). **g**, Side-by-side structural comparison of the USP14-bound RP in all states from the top view showing differential rotation of the lid and RPN1. **h**, Plots of the distance of pore-1 loop to the CP for those states not shown in Fig. 3c. The comparison shows that the pore-loop staircase architecture in state \(\{{\rm{S}}\}\}_{\{{\rm{B}}\}}^{\{{\rm{USP14}}\}}\}), \(\{{\rm{S}}\}\}_{\{{\rm{C}}\}}^{\{{\rm{USP14}}\}}\}) or \(\{{\rm{S}}\}\}_{\{{\rm{D4}}\}}^{\{{\rm{USP14}}\}}\}) is similar to that of E_A-like state. **i**, Root-mean-squared-deviation (RMSD) values of the ATPase structures are mapped between any pairs of the thirteen states. **j**, An integrated schematic diagram of proteasome state transitions illustrates the full functional cycles of the proteasome in the presence and absence of USP14. The solid circles are the states observed in the current study, whereas the dashed circles are the states observed in previous studies of substrate-free⁸ (orange) or substrate-engaged human proteasome¹⁰ (green) in the absence of USP14. Color blue and salmon label the substrate-

engaged and substrate-inhibited USP14-proteasome states, respectively. The states with closed and open CP gate are placed in pink and limon backgrounds, respectively. Vertical orange dashed lines link the state pairs with comparable AAA-ATPase structures. Black solid arrows and dashed arrows represent the possible structural transitions connecting the states observed in the current study and previous USP14-free studies^{8,10}, respectively

Source data

Extended Data Fig. 8 Dynamics of USP14 and key ATPase subunits in the proteasome.

a, Superposition of the USP14-RPT1 subcomplex structures from different states aligned against the RPT1 large AAA subdomain. USP14 rotates together with the RPT1 OB domain and moves up over 37 Å (from \{\{\rm{E}\}\}_{{\{\rm{D4}\}}}\^{{\{\rm{USP14}\}}}\) to \{\{\rm{E}\}\}_{{\{\rm{D0}\}}}\^{{\{\rm{USP14}\}}}\) relative to the RPT1 AAA domain. The angle between the OB domain and the AAA domain is measured and labelled for each state. **b**, Superposition of the USP14 structures from different states aligned against their USP domain. The UBL domain moves up over 24 Å (from \{\{\rm{E}\}\}_{{\{\rm{D4}\}}}\^{{\{\rm{USP14}\}}}\) to \{\{\rm{E}\}\}_{{\{\rm{D0}\}}}\^{{\{\rm{USP14}\}}}\) relative to the USP domain. The distance between Ser74 and Cys114 is measured and labelled for each state. **c**, Superposition of the RPT6 AAA domain structures from different states aligned against the large AAA subdomain. Left, comparison of the RPT6 AAA structures in the ATP-bound states. Middle, comparison of states \(\{\{\rm{S}\}\}_{{\{\rm{D4}\}}}\^{{\{\rm{USP14}\}}}\), \(\{\{\rm{E}\}\}_{{\{\rm{D2}\}}.1}\^{{\{\rm{USP14}\}}}\) and \(\{\{\rm{E}\}\}_{{\{\rm{D4}\}}}\^{{\{\rm{USP14}\}}}\), the ADP-bound states and state E_B (PDB ID: 6MSE) shows conformational changes of the AAA domain driven by the ATP hydrolysis and nucleotide exchange. Right, the open-gate states \(\{\{\rm{E}\}\}_{{\{\rm{D2}\}}.1}\^{{\{\rm{USP14}\}}}\), \(\{\{\rm{E}\}\}_{{\{\rm{D4}\}}}\^{{\{\rm{USP14}\}}}\), and \(\{\{\rm{S}\}\}_{{\{\rm{D4}\}}}\^{{\{\rm{USP14}\}}}\) show different refolding of both the pore-1 and pore-2 loops. **d**, Superposition of the RPT1 AAA

domain structures from different states aligned against the large AAA subdomain. Left, comparison of the structures in the ATP-bound states. Middle, comparison of the structures in different nucleotide-binding states. Right, the C-terminal tails of RPT6 exhibit three major orientations.

Extended Data Fig. 9 Structure-based site-directed mutagenesis.

a, Mapping of the potential RPT-binding sites and other residues affecting USP14 activation onto the USP14 structure in the $\langle\{\backslash rm\{E\}\}\rangle_{\langle\{\backslash rm\{D4\}\}\rangle^{\langle\{\backslash rm\{USP14\}\}\rangle}}$ model. The UBL domain of USP14 is not shown. **b**, Purification of USP14 mutants and analyzed by SDS/PAGE and stained with Coomassie blue. **c**, Peptidase activity assay was used to evaluate the effects of USP14 variants on regulating the CP gate opening. *P* values were analyzed using a two-tailed unpaired *t*-test between wild-type USP14-bound and USP14-free proteasomes. The results suggest that the mutants promote the CP gate opening to the same degree as that of wild-type USP14 as compared to that of the USP14-free proteasome. **d** and **e**, Ubiquitin-AMC hydrolysis assay to measure the DUB activity of USP14 mutants in the presence (panel **d**) or absence (panel **e**) of the human proteasome. Data are presented as mean \pm s.d. from three independent experiments. *P* values were analyzed using a two-tailed unpaired *t*-test between USP14 mutants and wild-type USP14. *P* value is not labelled for data with *P* $>$ 0.05, which is not significant. Data in **c–e** are presented as mean \pm s.d. from three independent experiments, each with three replicates. Dots, individual data points. **f**, The DUB activity of USP14 mutants in the presence or absence of proteasome. Data points are the average of individual data points shown in panels (**d**) and (**e**) and Fig. 2*i*. **g** and **h**, *In vitro* degradation of Ub_n -Sic1^{PY} by the 26S proteasome in the presence of USP14 mutants testing the USP-OB interface or linker region (panel **g**) and affecting the USP-AAA interfaces (panel **h**) (repeated three times with similar results). Samples were analyzed by SDS-PAGE/Western blot using anti-T7 antibody. TEEQ, insertion of TEEQ after residue 92 in the linker region. Δ93–96, mutant with deletion of residues 93–96 in the linker region. W/O, the proteasome without binding to USP14. WT, the wildtype USP14-bound proteasome. **i**, Multiple sequence alignment of USP14 from five species was performed by Chimera. Annotation is based on the structural

and mutational data from Figs. 2 and 3. The mutations with the strongest phenotypes (marked by red stars) all correspond to the amino acids (highlighted bold) that are fully conserved from yeast to human. Those mutations with moderate phenotypes correspond to the amino acids that are well conserved in mammals but may vary in yeast. For gel source data, see Supplementary Fig. 1

[Source data](#)

Extended Data Table 1 Cryo-EM data collection, refinement and validation statistics

Extended Data Table 2 Summary of key structural features of USP14-bound proteasome states

Supplementary information

[Supplementary Figure 1](#)

This file contains the raw (uncropped) gel images for Fig. 2 and for Extended Data Figs. 1, 9.

[Reporting Summary](#)

[Peer Review File](#)

Source data

[Source Data Fig. 2](#)

[Source Data Fig. 3](#)

[Source Data Extended Data Fig. 1](#)

[Source Data Extended Data Fig. 7](#)

Source Data Extended Data Fig. 9

Rights and permissions

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

Reprints and Permissions

About this article

Cite this article

Zhang, S., Zou, S., Yin, D. *et al.* USP14-regulated allosteric of the human proteasome by time-resolved cryo-EM. *Nature* **605**, 567–574 (2022).
<https://doi.org/10.1038/s41586-022-04671-8>

- Received: 24 August 2021
- Accepted: 22 March 2022
- Published: 27 April 2022
- Issue Date: 19 May 2022
- DOI: <https://doi.org/10.1038/s41586-022-04671-8>

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

[Control of human protein-degradation machinery revealed](#)

Nature Research Briefing 27 Apr 2022

This article was downloaded by **calibre** from <https://www.nature.com/articles/s41586-022-04671-8>

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

Amendments & Corrections

- [**Author Correction: The guidance receptor plexin D1 is a mechanosensor in endothelial cells**](#) [03 May 2022]
Author Correction •
- [**Publisher Correction: Signatures of TOP1 transcription-associated mutagenesis in cancer and germline**](#) [03 May 2022]
Publisher Correction •
- [**Addendum: Large metallicity variations in the Galactic interstellar medium**](#) [03 May 2022]
Addendum •

[Download PDF](#)

- Author Correction
- [Published: 03 May 2022](#)

Author Correction: The guidance receptor plexin D1 is a mechanosensor in endothelial cells

- [Vedanta Mehta^{1,2},](#)
- [Kar-Lai Pang^{1,2},](#)
- [Daniel Rozbesky^{2,3},](#)
- [Katrín Náther^{1,2},](#)
- [Adam Keen^{1,2},](#)
- [Dariusz Lachowski⁴,](#)
- [Youxin Kong^{2,3},](#)
- [Dimple Karia^{2,3},](#)
- [Michael Ameismeier^{2,3},](#)
- [Jianhua Huang⁵,](#)
- [Yun Fang⁶,](#)
- [Armando del Rio Hernandez⁴,](#)
- [John S. Reader^{1,2 na1},](#)
- [E. Yvonne Jones^{2,3 na1} &](#)
- [Ellie Tzima^{1,2 na1}](#)

[Nature](#) volume 605, page E6 (2022)

- 399 Accesses
- 1 Altmetric
- [Metrics details](#)

Subjects

- [Cardiovascular biology](#)
- [Mechanisms of disease](#)

The [Original Article](#) was published on 05 February 2020

Correction to: *Nature* <https://doi.org/10.1038/s41586-020-1979-4> Published online 05 February 2020

In this Article, the blots in Extended Data Figs 9 and 11 were inadvertently taken from different replicates of the same experiment. In the Supplementary Information to this Amendment, we have replaced these blots with the corresponding replicate experiment, shown together with the original figures for comparison. In addition, Supplementary Fig. 1 of the original Article has been revised to include the corresponding uncropped western blot images, including the correct matching blots shown in the original Figs 2a (ERK1/2), 3b (PI3K/P85 and VE-cadherin), 3d, 4h (pAKT, pERK1/2 and pVEGFR2), 4i (pERK1/2), and Extended Data Figs 1b (PlxnD1), 1c (GAPDH), 1d (NRP1 and GAPDH), 1f (G α q/11 and GAPDH), 1g (GAPDH), 3a (p-eNOS, pAkt and pERK1/2), 7d (VEGFR2), and 8. Replicates for all blots are provided. These changes do not affect the results or conclusions of the Article, and we apologise for any inconvenience these errors may have caused. The original Article has not been corrected online

Author information

Author notes

1. These authors jointly supervised this work: John S. Reader, E. Yvonne Jones, Ellie Tzima

Affiliations

1. Cardiovascular Medicine, Radcliffe Department of Medicine,
University of Oxford, Oxford, UK

Vedanta Mehta, Kar-Lai Pang, Katrin Nather, Adam Keen, John S.
Reader & Ellie Tzima

2. Wellcome Centre for Human Genetics, University of Oxford, Oxford,
UK

Vedanta Mehta, Kar-Lai Pang, Daniel Rozbesky, Katrin Nather, Adam
Keen, Youxin Kong, Dimple Karia, Michael Ameismeier, John S.
Reader, E. Yvonne Jones & Ellie Tzima

3. Division of Structural Biology, Wellcome Centre for Human Genetics,
University of Oxford, Oxford, UK

Daniel Rozbesky, Youxin Kong, Dimple Karia, Michael
Ameismeier & E. Yvonne Jones

4. Cellular and Molecular Biomechanics Laboratory, Department of
Bioengineering, Imperial College London, London, UK

Dariusz Lachowski & Armando del Rio Hernandez

5. Department of Medicine, Duke University, Durham, NC, USA

Jianhua Huang

6. Department of Medicine, University of Chicago, Chicago, IL, USA

Yun Fang

Corresponding author

Correspondence to [Ellie Tzima](#).

Additional information

Supplementary information is available in the online version of this Amendment.

Supplementary information

Supplementary Figure 1

This file shows the original and corrected Extended Data Figs 9 and 11.

Supplementary Figure 2

This file shows the corrected Supplementary Fig. 1 of the original Article, containing uncropped western blots for all figures.

Supplementary Figure 3

This file shows replicates of all western blots.

Rights and permissions

Reprints and Permissions

About this article

Cite this article

Mehta, V., Pang, KL., Rozbesky, D. *et al.* Author Correction: The guidance receptor plexin D1 is a mechanosensor in endothelial cells. *Nature* **605**, E6 (2022). <https://doi.org/10.1038/s41586-022-04815-w>

- Published: 03 May 2022
- Issue Date: 19 May 2022
- DOI: <https://doi.org/10.1038/s41586-022-04815-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/s41586-022-04815-w>

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

[Download PDF](#)

- Publisher Correction
- Open Access
- [Published: 03 May 2022](#)

Publisher Correction: Signatures of TOP1 transcription-associated mutagenesis in cancer and germline

- [Martin A. M. Reijns](#) [ORCID: orcid.org/0000-0002-5048-2752](#)¹_{na1},
- [David A. Parry](#) [ORCID: orcid.org/0000-0003-0376-7736](#)¹_{na1},
- [Thomas C. Williams](#) [ORCID: orcid.org/0000-0002-3866-1344](#)^{1,2}_{na1},
- [Ferran Nadeu](#) [ORCID: orcid.org/0000-0003-2910-9440](#)^{3,4},
- [Rebecca L. Hindshaw](#)⁵,
- [Diana O. Rios Szwed](#) [ORCID: orcid.org/0000-0003-0893-6324](#)¹,
- [Michael D. Nicholson](#) [ORCID: orcid.org/0000-0003-2813-8979](#)⁶,
- [Paula Carroll](#)¹,
- [Shelagh Boyle](#)⁷,
- [Romina Royo](#)⁸,
- [Alex J. Cornish](#) [ORCID: orcid.org/0000-0002-3966-3501](#)⁹,
- [Hang Xiang](#)¹⁰,
- [Kate Ridout](#)¹¹,
- [The Genomics England Research Consortium](#),
- [Colorectal Cancer Domain UK 100,000 Genomes Project](#),
- [Anna Schuh](#) [ORCID: orcid.org/0000-0002-3938-8490](#)¹¹,
- [Konrad Aden](#)¹⁰,
- [Claire Palles](#) [ORCID: orcid.org/0000-0002-9670-2263](#)⁵,
- [Elias Campo](#) [ORCID: orcid.org/0000-0001-9850-9793](#)^{3,4,12,13},

- [Tatjana Stankovic](#) ORCID: [orcid.org/0000-0002-3780-274X⁵](https://orcid.org/0000-0002-3780-274X),
- [Martin S. Taylor](#) ORCID: [orcid.org/0000-0001-7656-330X²](https://orcid.org/0000-0001-7656-330X) &
- [Andrew P. Jackson](#) ORCID: [orcid.org/0000-0002-8739-2646¹](https://orcid.org/0000-0002-8739-2646)

Nature volume **605**, page E7 (2022)

- 304 Accesses
- 2 Altmetric
- [Metrics details](#)

Subjects

- [Cancer genomics](#)
- [Genomic instability](#)
- [Nucleotide excision repair](#)

The [Original Article](#) was published on 09 February 2022

Correction to: *Nature* <https://doi.org/10.1038/s41586-022-04403-y>
Published online 9 February 2022

In the version of this article initially published, there were errors in the affiliations shown for Colorectal Cancer Domain UK 100,000 Genomes Project members Steve Thorn and Henry Wood. Steve Thorn is with the Cancer Research UK Edinburgh Centre, Institute of Genetics and Cancer, The University of Edinburgh, Edinburgh, UK (not University of Leeds). Henry Wood is with Pathology and Data Analytics, Leeds Institute of Medical Research, St James's University Hospital, University of Leeds, Leeds, UK (not The University of Edinburgh). The errors have been corrected in the HTML and PDF versions of the article.

Author information

Author notes

1. These authors contributed equally: Martin A. M. Reijns, David A. Parry, Thomas C. Williams

Affiliations

1. Disease Mechanisms, MRC Human Genetics Unit, Institute of Genetics and Cancer, The University of Edinburgh, Edinburgh, UK

Martin A. M. Reijns, David A. Parry, Thomas C. Williams, Diana O. Rios Szwed, Paula Carroll & Andrew P. Jackson

2. Biomedical Genomics, MRC Human Genetics Unit, Institute of Genetics and Cancer, The University of Edinburgh, Edinburgh, UK

Thomas C. Williams & Martin S. Taylor

3. Institut d'Investigacions Biomèdiques August Pi i Sunyer (IDIBAPS), Barcelona, Spain

Ferran Nadeu & Elias Campo

4. Centro de Investigación Biomédica en Red de Cáncer (CIBERONC), Madrid, Spain

Ferran Nadeu & Elias Campo

5. Institute of Cancer and Genomic Sciences, University of Birmingham, Birmingham, UK

Rebecca L. Hindshaw, Boris Noyvert, Claire Palles & Tatjana Stankovic

6. Cancer Research UK Edinburgh Centre, Institute of Genetics and Cancer, The University of Edinburgh, Edinburgh, UK

Michael D. Nicholson, Steve Thorn & Ian Tomlinson

7. Genome Regulation, MRC Human Genetics Unit, Institute of Genetics and Cancer, The University of Edinburgh, Edinburgh, UK

Shelagh Boyle

8. Barcelona Supercomputing Center (BSC), Barcelona, Spain

Romina Royo

9. The Institute of Cancer Research, London, UK

Alex J. Cornish, Daniel Chubb, Alex Cornish, Ben Kinnersley, Richard Houlston, Andrea Sottoriva, Giulio Caravagna & Richard Culliford

10. Institute of Clinical Molecular Biology, Christian-Albrechts-University and University Hospital Schleswig-Holstein, Kiel, Germany

Hang Xiang & Konrad Aden

11. Department of Oncology, University of Oxford, Oxford, UK

Kate Ridout & Anna Schuh

12. Hospital Clínic of Barcelona, Barcelona, Spain

Elias Campo

13. Departament de Fonaments Clínics, Universitat de Barcelona, Barcelona, Spain

Elias Campo

14. Genomics England, London, UK

John C. Ambrose, Prabhu Arumugam, Roel Bevers, Marta Bleda, Freya Boardman-Pretty, Christopher R. Bousted, Helen Brittain, Mark J. Caulfield, Georgia C. Chan, Greg Elgar, Tom Fowler, Adam Giess, Angela Hamblin, Shirley Henderson, Tim J. P. Hubbard, Rob Jackson, Louise J. Jones, Dalia Kasperaviciute, Melis

Kayikci, Athanasios Kousathanas, Lea Lahnstein, Sarah E. A. Leigh, Ivonne U. S. Leong, Javier F. Lopez, Fiona Maleady-Crowe, Meriel McEntagart, Federico Minneci, Loukas Moutsianas, Michael Mueller, Nirupa Murugaesu, Anna C. Need, Peter O'Donovan, Chris A. Odhams, Christine Patch, Mariana Buongermino Pereira, Daniel Perez-Gil, John Pullinger, Tahrima Rahim, Augusto Rendon, Tim Rogers, Kevin Savage, Kushmita Sawant, Richard H. Scott, Afshan Siddiq, Alexander Sieghart, Samuel C. Smith, Alona Sosinsky, Alexander Stuckey, Mélanie Tanguy, Ana Lisa Taylor Tavares, Ellen R. A. Thomas, Simon R. Thompson, Arianna Tucci, Matthew J. Welland, Eleanor Williams, Katarzyna Witkowska & Suzanne M. Wood

15. William Harvey Research Institute, Queen Mary University of London, London, UK

Freya Boardman-Pretty, Mark J. Caulfield, Greg Elgar, Shirley Henderson, Louise J. Jones, Dalia Kasperaviciute, Loukas Moutsianas, Michael Mueller, Anna C. Need, Christine Patch, Alona Sosinsky, Ellen R. A. Thomas, Arianna Tucci, Katarzyna Witkowska & Suzanne M. Wood

16. Manchester Interdisciplinary Biocentre, University of Manchester, Manchester, UK

David Wedge & Andreas Gruber

17. Wellcome Centre for Human Genetics, Oxford, UK

Anna Frangou & David Church

18. Cancer Institute, University College London, London, UK

William Cross

19. Barts Cancer Institute, Barts and The London School of Medicine and Dentistry, Queen Mary University of London, London, UK

Trevor Graham

20. Institute for Research in Biomedicine (IRB Barcelona), The Barcelona Institute of Science and Technology (BIST), Barcelona, Spain

Nuria Lopez-Bigas & Claudia Arnedo-Pac

21. Pathology and Data Analytics, Leeds Institute of Medical Research, St James's University Hospital, University of Leeds, Leeds, UK

Phil Quirke & Henry Wood

Consortia

The Genomics England Research Consortium

- , John C. Ambrose
- , Prabhu Arumugam
- , Roel Bevers
- , Marta Bleda
- , Freya Boardman-Pretty
- , Christopher R. Boustred
- , Helen Brittain
- , Mark J. Caulfield
- , Georgia C. Chan
- , Greg Elgar
- , Tom Fowler
- , Adam Giess
- , Angela Hamblin
- , Shirley Henderson
- , Tim J. P. Hubbard
- , Rob Jackson
- , Louise J. Jones
- , Dalia Kasperaviciute
- , Melis Kayikci
- , Athanasios Kousathanas
- , Lea Lahnstein

- , Sarah E. A. Leigh
- , Ivonne U. S. Leong
- , Javier F. Lopez
- , Fiona Maleady-Crowe
- , Meriel McEntagart
- , Federico Minneci
- , Loukas Moutsianas
- , Michael Mueller
- , Nirupa Murugaesu
- , Anna C. Need
- , Peter O'Donovan
- , Chris A. Odhams
- , Christine Patch
- , Mariana Buongermino Pereira
- , Daniel Perez-Gil
- , John Pullinger
- , Tahrima Rahim
- , Augusto Rendon
- , Tim Rogers
- , Kevin Savage
- , Kushmita Sawant
- , Richard H. Scott
- , Afshan Siddiq
- , Alexander Sieghart
- , Samuel C. Smith
- , Alona Sosinsky
- , Alexander Stuckey
- , Mélanie Tanguy
- , Ana Lisa Taylor Tavares
- , Ellen R. A. Thomas
- , Simon R. Thompson
- , Arianna Tucci
- , Matthew J. Welland
- , Eleanor Williams
- , Katarzyna Witkowska
- & Suzanne M. Wood

Colorectal Cancer Domain UK 100,000 Genomes Project

- Daniel Chubb
- , Alex Cornish
- , Ben Kinnersley
- , Richard Houlston
- , David Wedge
- , Andreas Gruber
- , Anna Frangou
- , William Cross
- , Trevor Graham
- , Andrea Sottoriva
- , Giulio Caravagna
- , Nuria Lopez-Bigas
- , Claudia Arnedo-Pac
- , David Church
- , Richard Culliford
- , Steve Thorn
- , Phil Quirke
- , Henry Wood
- , Ian Tomlinson
- & Boris Noyvert

Corresponding authors

Correspondence to [Martin A. M. Reijns](#), [Martin S. Taylor](#) or [Andrew P. Jackson](#).

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

Reijns, M.A.M., Parry, D.A., Williams, T.C. *et al.* Publisher Correction: Signatures of TOP1 transcription-associated mutagenesis in cancer and germline. *Nature* **605**, E7 (2022). <https://doi.org/10.1038/s41586-022-04812-z>

- Published: 03 May 2022
- Issue Date: 19 May 2022
- DOI: <https://doi.org/10.1038/s41586-022-04812-z>

Share this article

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

[Get shareable link](#)

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

[Copy to clipboard](#)

Provided by the Springer Nature SharedIt content-sharing initiative

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

- Addendum
- [Published: 03 May 2022](#)

Addendum: Large metallicity variations in the Galactic interstellar medium

- [Annalisa De Cia](#) [ORCID: orcid.org/0000-0003-2082-1626¹](#),
- [Edward B. Jenkins](#) [ORCID: orcid.org/0000-0003-1892-4423²](#),
- [Andrew J. Fox](#) [ORCID: orcid.org/0000-0003-0724-4115³](#),
- [Cédric Ledoux](#) [ORCID: orcid.org/0000-0002-7864-3327⁴](#),
- [Tanita Ramburuth-Hurt](#) [ORCID: orcid.org/0000-0003-2597-2415¹](#),
- [Christina Konstantopoulou](#) [ORCID: orcid.org/0000-0002-4690-0157¹](#),
- [Patrick Petitjean⁵](#) &
- [Jens-Kristian Krogager](#) [ORCID: orcid.org/0000-0002-4912-9388¹](#)

[Nature](#) volume 605, page E8 (2022)

- 287 Accesses
- 1 Altmetric
- [Metrics details](#)

Subjects

- [Astrophysical dust](#)
- [Galaxies and clusters](#)
- [Interstellar medium](#)

The [Original Article](#) was published on 08 September 2021

Addendum to: *Nature* <https://doi.org/10.1038/s41586-021-03780-0>

Published online 08 September 2021, corrected 03 November 2021

In this Article, we stated that significant amounts of low-metallicity gas in the neutral interstellar medium (ISM) are needed to explain the observed abundance patterns. We clarify here that the exact amount of low-metallicity gas present along each line of sight is difficult to quantify because several assumptions are required to derive the gas mass from the column densities measured along the line of sight. The volatile elements often deviate from the linear fit to the more refractory elements in the abundance patterns (Extended Data Fig. 3 of the original Article). As suggested in the Article, this indicates that there must be a mixture of different gas types along many lines of sight. Near-solar metallicity gas could give rise to the volatile elements, while the low-metallicity nearly dust-free gas could dominate the abundance patterns of the refractory elements. The exact composition of this gas mixture is hard to determine. Our results favour the possibility that an amount between a few per cent and almost half of the gas has a low metallicity, but we cannot rule out a mixture of solar-metallicity gas having vastly different levels of depletion onto dust grains. These results do not contradict the observations that HII regions and OB stars show smaller scatter in metallicity, particularly if the mass contribution of the low-metallicity gas is small, and in general given that some of the neutral gas is in an extended phase of the ISM rather than tracing denser (and possibly more mixed) star-forming environments. Our results do show, however, that low-metallicity gas is present in the ISM, the mass of which is unconstrained so far. Our observations highlight the variety in chemical enrichment of the neutral ISM, both in terms of metallicity and dust depletion. A more complete analysis of the chemical properties of the ISM is under way

Author information

Affiliations

1. Department of Astronomy, University of Geneva, Versoix, Switzerland

Annalisa De Cia, Tanita Ramburuth-Hurt, Christina Konstantopoulou & Jens-Kristian Krogager

2. Princeton University Observatory, Princeton, NJ, USA

Edward B. Jenkins

3. AURA for ESA, Space Telescope Science Institute, Baltimore, MD, USA

Andrew J. Fox

4. European Southern Observatory, Santiago, Chile

Cédric Ledoux

5. Institut d’Astrophysique de Paris, Sorbonne Universités & CNRS, Paris, France

Patrick Petitjean

Corresponding author

Correspondence to [Annalisa De Cia](#).

Rights and permissions

[Reprints and Permissions](#)

About this article

Cite this article

De Cia, A., Jenkins, E.B., Fox, A.J. *et al.* Addendum: Large metallicity variations in the Galactic interstellar medium. *Nature* **605**, E8 (2022). <https://doi.org/10.1038/s41586-022-04811-0>

- Published: 03 May 2022
- Issue Date: 19 May 2022
- DOI: <https://doi.org/10.1038/s41586-022-04811-0>

Share this article

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

[Get shareable link](#)

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

[Copy to clipboard](#)

Provided by the Springer Nature SharedIt content-sharing initiative

This article was downloaded by **calibre** from <https://www.nature.com/articles/s41586-022-04811-0>

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

Innovations In