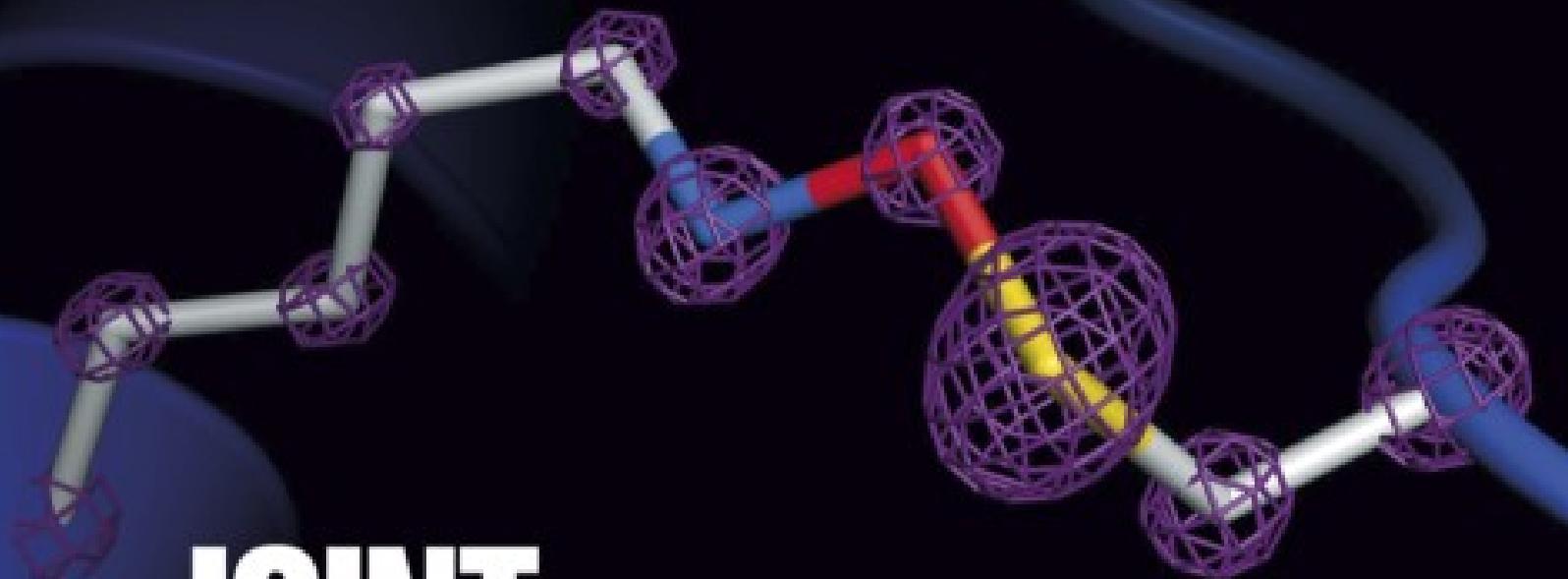


# nature



## JOINT VENTURE

Nitrogen, oxygen and sulfur connect to form an unexpected crosslink in protein structures

**Coronavirus**  
COVID-19 offers unique insights into human behaviour

**Regulate research**  
Time for a law to protect the remains of Black Americans

**Zombie fires**  
The effect on boreal forests of fires that survive the winter

# Nature.2021.05.22

[Sat, 22 May 2021]

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

---

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

---

# Donation

---

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

# This Week

- **[Tackling systemic racism requires the system of science to change](#)** [ 19 May 2021]  
Editorial • In response to the global Black Lives Matter protests, many institutions pledged actions to combat racism. That's not enough.
- **[Universal health care must be a priority — even amid COVID](#)** [ 18 May 2021]  
Editorial • A focus on specific diseases has derailed efforts to achieve health care for all before. The world must not repeat that mistake with COVID-19.
- **[Media bias delegitimizes Black-rights protesters](#)** [ 19 May 2021]  
World View • Linguistic analyses show that powerful sources and sensationalist terms have long dominated coverage of civil-rights protests.
- **[Microbes in Neanderthals' mouths reveal their carb-laden diet](#)** [ 14 May 2021]  
Research Highlight • Gunk on ancient teeth yields bacterial DNA, allowing scientists to trace the oral microbiome's evolution.
- **[Holding a tool wrong? This brain region will notice](#)** [ 13 May 2021]  
Research Highlight • The part of the brain that assesses tool usage is normally more focused on hands than on tools.
- **[How to tell the thermodynamic cost of time](#)** [ 13 May 2021]  
Research Highlight • Quantum and ordinary clocks alike gain accuracy as they give off more heat.
- **[Voyager 1 captures faint ripples in the stuff between the stars](#)** [ 13 May 2021]  
Research Highlight • The first spacecraft to visit interstellar space has now become the first to make continuous measurements of waves in that remote realm.
- **[Flashy plants draw outsize share of scientists' attention](#)** [ 11 May 2021]  
Research Highlight • Blue-flowered plants get the most scientific love; those with green or brown flowers, not so much.

- **Kids with ‘bubble baby disease’ see gains after gene therapy** [ 14 May 2021]

Research Highlight • Stem cells carrying corrective genes offer hope for a devastating immune disorder.

- **A deep look into the physics of earthquake slip** [ 13 May 2021]

Research Highlight • Laboratory experiments provide a glimpse of what happens when rocks fail.

## EDITORIAL

19 May 2021

# Tackling systemic racism requires the system of science to change

In response to the global Black Lives Matter protests, many institutions pledged actions to combat racism. That's not enough.



The Black Lives Matter protests prompted many institutions to acknowledge and confront racial injustice. Credit: Linnea Rheborg/Getty

Next week marks a year since the murder of George Floyd, and nearly a year since the Black Lives Matter protests compelled numerous institutions — including many in research — to acknowledge systemic racism. These events made universities, institutes, corporations, museums, societies,

publishers and funders confront racial injustice in a way that had never happened before.

As part of that response, *Nature* recognized systemic racism in science and our part in it, and [committed to stand against it](#).

We know that such statements must be followed by actions. At *Nature*, we have made it an editorial priority to expose and tackle racism in science by publishing more research, commentary and journalism about racism and racial injustice. Next year, we will produce a special issue, under the guidance of a group of external editors, that examines systemic racism in research. We will be launching a news internship for Black journalists later this year. We are taking further steps to diversify our authors, reviewers and contributors. And we know that too few of our editorial staff are people of colour, so we are working to change this.



### [Major physics society won't meet in cities with racist policing record](#)

The other journals and teams in the *Nature* Portfolio are also forging stronger connections with communities of Black researchers; and our publisher, Springer Nature, has made commitments to champion diversity internally and in the communities it serves. Its Black Employee Network, formed in August 2020, has made valuable contributions to editorial policy and to elevating Black voices in science, technology, engineering and mathematics (STEM). Many other journals and science publishers have also made welcome changes.

But we know we are only at the foothills; there is a mountain ahead. We need to do much more, and are determined to do so. At the same time, we recognize that such pledges and actions, by themselves, do not constitute systemic change.

Racism in science is endemic because the systems that produce and teach scientific knowledge have, for centuries, misrepresented, marginalized and mistreated people of colour and under-represented communities. The research system has justified racism — and, too often, scientists in positions of power have benefited from it. That system includes the organization of research: how it is funded, published and evaluated.



### [What the data say about police brutality and racial bias — and which reforms might work](#)

Ending systemic racism will therefore require those in the system, including *Nature*, to collectively acknowledge and study these facts, and to ask: how and why did this happen? We need to thoroughly understand the root causes, even as we seek energetically to remedy the ongoing damage. Some have already started down this road. Projects at the Massachusetts Institute of Technology in Cambridge and the University of Glasgow, UK, have investigated these institutions' past ties to the slave trade and how they prospered from it, helping to build a more accurate and complete account of science history.

Hundreds of individual organizations have pledged actions to combat racism. All of these are important, but on their own they will not bring about the systemic change that is required. One essential change all institutions can make today is to put the right incentives in place. They must ensure that anti-racism is embedded in their organization's objectives and that such work wins recognition and promotion. Too often, conventional metrics — citations, publication, profits — reward those in positions of power, rather than helping to shift the balance of power.

A second change institutions should make is to come together to tackle racism, as some already are. At the very least, this means talking to and learning from a wide range of communities, and transcending conventional boundaries to team up. Funders, research institutions and publishers must work together to ensure that research from diverse scientists is funded and published. As part of the system of science, *Nature* is starting to develop such partnerships, and we look forward to doing more. Together, we will move further, faster.

Nature **593**, 313 (2021)

doi: <https://doi.org/10.1038/d41586-021-01312-4>

## Jobs from Nature Careers

- - - [All jobs](#)
    - 
    - [Post-Doc - Numerical and microfabrication development of 4D metamaterials for mechanical, acoustic and photonics applications](#)

[Université Bourgogne Franche-Comté \(UBFC\)](#)

[BESANCON, France](#)

JOB POST

- **Postdoctoral Position - Ecological Modeler**

The University of British Columbia (UBC)

Vancouver, Canada

JOB POST

- **Postdoctoral Fellow | Zandstra Stem Cell Bioengineering Lab**

The University of British Columbia (UBC)

Vancouver, Canada

JOB POST

- **Personalreferent\*in (d/m/w), Gruppe Personaladministration**

Helmholtz Centre for Heavy Ion Research GmbH (GSI)

Darmstadt, Germany

JOB POST

Download PDF

Download PDF

## Related Articles



- [Major physics society won't meet in cities with racist policing record](#)



- [What the data say about police brutality and racial bias — and which reforms might work](#)



- [Grieving and frustrated: Black scientists call out racism in the wake of police killings](#)



- [Systemic racism: science must listen, learn and change](#)

---

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-021-01312-4>

## EDITORIAL

18 May 2021

# Universal health care must be a priority — even amid COVID

A focus on specific diseases has derailed efforts to achieve health care for all before. The world must not repeat that mistake with COVID-19.



The COVID pandemic illustrated how a lack of reliable health care renders communities vulnerable. Credit: John Moore/Getty

Vaccinating the world's population against COVID-19 remains a global health priority. But it is vital that this effort does not overshadow the need to ensure that everyone, everywhere has access to basic health care.

Despite the urgency of the current crisis, the provision of universal health care remains a priority for Tedros Adhanom Ghebreyesus, the director-general of the World Health Organization (WHO). It is also enshrined in the United Nations Sustainable Development Goals on the basis that health is a prerequisite for economic growth. Governments, scientists and the public should support this goal, because it's in everyone's best interests. And they will have an opportunity next week, when the World Health Assembly convenes online.

The pandemic has amply demonstrated how a lack of reliable health care can render communities vulnerable. Although access to health care isn't the sole factor that determined how well countries fared, its absence clearly fuelled the flames. Many lives have been lost in India because hospitals have been overwhelmed. In the United States, COVID-19 deaths have been higher among people on low incomes, who are less likely to have health insurance and therefore less likely to seek medical care promptly. A similar pattern has been seen elsewhere: one study found that in the poorer neighbourhoods of Santiago, more than 90% of people whose deaths were attributed to COVID-19 died outside health-care facilities ([G. E. Mena et al. \*Science\* https://doi.org/f9b4; 2021](https://doi.org/f9b4)). Moreover, people without reliable health care might be more vulnerable to complications of COVID-19 because of poorly controlled chronic diseases.



[Will COVID force public health to confront America's epic inequality?](#)

A lack of easily accessible health care — and of health systems for sharing information — has impeded the detection and monitoring of COVID-19 infections. Should another deadly virus emerge in a region with inadequate health care, the world could lose valuable time to contain the outbreak. The two largest Ebola outbreaks so far — in West Africa and the Democratic Republic of the Congo — spread for weeks to months before they were identified.

Despite almost a century of calls to provide all people with health care, attempts have been stymied by crises and disease-specific interventions. After the economic depression of the 1930s, international health officials working for the League of Nations touted the need for the provision of basic health services by country-wide networks of clinics. This vision was interrupted by the Second World War, but was revived in 1946, when the newly formed United Nations met to write a constitution for the WHO. The constitution stated that health is a human right, and that governments are responsible for the health of their people.

But the United States nearly rejected the WHO and its constitution outright. At the time, opponents of national health-care provision in the country connected the measure to socialism and communism. The United States eventually signed on, but inserted a clause stating that it could withdraw from the WHO at any time — meaning that the country donating the most money (dues are based on the size of economies) could walk away if it opposed the organization's ideology.



### How to defuse malaria's ticking time bomb

The next 20 years of single-disease programmes driven by the WHO and global health funders wasn't ideological, however. Ironically, they were also driven by a wave of scientific advances that offered simple, technological fixes to specific health problems, such as the use of the insecticide DDT to fight malaria and antibiotics to fight infections.

But, in 1978, the push to build up health systems was revived at an international conference on providing everyone with primary health care, held in Alma-Ata in what was then the Soviet Union. The resulting WHO-sponsored Alma-Ata declaration vowed to provide essential care, at the level of neighbourhoods, by the year 2000.

But according to Tedros and health-policy experts, the Alma-Ata declaration was undermined by factors including inadequate political leadership, economic crises, political instability and an over-investment in treating individual diseases ([T. A. Ghebreyesus et al. Lancet 392, 1371–1372; 2018](#)).

Others have suggested that the movement lacked defined steps backed by evidence, as well as cost-effectiveness assessments. Compare this with the UN children's charity, UNICEF, which in the 1980s vowed to save the lives of four million to five million children a year through well-defined and budgeted programmes to deliver vaccines for diseases including measles, tetanus and polio. Government and philanthropic donors grasped the concept

immediately, and UNICEF quickly became one of the larger UN agencies. In 2019, its budget was nearly three times that of the WHO.

Childhood immunization programmes save lives, but the lack of investment in strengthening countries' health systems has led to untold deaths. The answer isn't to stop vaccinations, of course, but to take cues from the success of UNICEF's campaign and the failures of Alma-Ata.



### [Why did the world's pandemic warning system fail when COVID hit?](#)

In 2019, the WHO once again turned the focus on health care for all, this time at the first UN high-level meeting on universal health care. A corresponding report stated that to provide all people with primary health care, countries, on average, must increase their spending in this area by 1% of their gross domestic product. And world leaders signed a declaration promising to pursue universal health care — in their national context — and provide basic, affordable health services to everyone in the country. To hold them accountable, global-health researchers have created an online portal to track progress towards the attainment of this goal by 2030. For example, the tracker says that about 15% of the populations of the United States and Cuba lack access to essential health services. The rate grows to 20% in China and 45% in India and Kenya.

The WHO has placed 'health for all' high on the agenda of next week's meeting, hoping to drive political and financial commitments from governments. Perhaps mindful of the vagueness that doomed past efforts,

Tedros has created a new council of economists, health and development experts to advise on the economics of providing everyone with basic health care, including ways to quantify its value.

Universal health care might seem a lofty goal amid a crisis, but if we do not push for change now, we will regret it. The pandemic has increased the number of people living in extreme poverty, making them more vulnerable to disease. It's infected, killed and traumatized health-care workers everywhere, most devastatingly in places that had too few already. "Our failure to invest in health systems doesn't only leave individuals, families and communities at risk, it also leaves the world vulnerable to outbreaks and other health emergencies," Tedros said in October 2019. "A pandemic could bring economies and nations to their knees." A few months later, it did. We must not let that happen again.

Nature **593**, 313-314 (2021)

doi: <https://doi.org/10.1038/d41586-021-01313-3>

## Jobs from Nature Careers

- - - [All jobs](#)
    - 
    - [Post-Doc - Numerical and microfabrication development of 4D metamaterials for mechanical, acoustic and photonics applications](#)
      - [Université Bourgogne Franche-Comté \(UBFC\)](#)
      - [BESANCON, France](#)
      - [JOB POST](#)

[The University of British Columbia \(UBC\)](#)

[Vancouver, Canada](#)

[JOB POST](#)

- [Postdoctoral Fellow | Zandstra Stem Cell Bioengineering Lab](#)

[The University of British Columbia \(UBC\)](#)

[Vancouver, Canada](#)

[JOB POST](#)

- [Personalreferent\\*in \(d/m/w\), Gruppe Personaladministration](#)

[Helmholtz Centre for Heavy Ion Research GmbH \(GSI\)](#)

[Darmstadt, Germany](#)

[JOB POST](#)

[Download PDF](#)

[Download PDF](#)

## Related Articles

-  [Why did the world's pandemic warning system fail when COVID hit?](#)

-  **Will COVID force public health to confront America's epic inequality?**
-  **How to defuse malaria's ticking time bomb**

---

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-021-01313-3>

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

WORLD VIEW

19 May 2021

# Media bias delegitimizes Black-rights protesters



Linguistic analyses show that powerful sources and sensationalist terms have long dominated coverage of civil-rights protests.

- [Danielle Kilgo](#) 0

1. [Danielle Kilgo](#)

1. Danielle Kilgo is a professor of journalism, diversity and equality at the University of Minnesota in Minneapolis.

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

The protests following the killing of George Floyd, an unarmed Black man, by police a year ago built on those that came before — in response to the deaths of Michael Brown, Eric Garner, Tamir Rice, Sandra Bland and far too many others. The global reckoning was a result of decades of work by advocates who prepared the public to engage with race and racism. One reason their message had taken so long to become mainstream lies in how the press typically covers protests.

I study media representation, marginalized communities and social movements. I have quantified narratives in news coverage of Black civil rights since the 2012 killing of Trayvon Martin, comparing it with coverage of protests for and against former US president Donald Trump, women's rights, gun control, environmental protection and more. My colleagues and I use computational methods to find linguistic patterns, rhetoric and sentiment in texts, together with human coding for overarching themes including 'violence', 'combativeness' and 'racial justice', as well as for contextual cues, such as the passive voice in headlines, for example "peaceful protesters teargassed", which neglect to say who took the action.

Linguistic analysis can show what narratives are being presented to and adopted by the public. Such work — examining which groups are privileged at the expense of others — can help many enterprises, including the scientific system, to repair damage from stigmatizing narratives.

Civil-rights protesters are the least likely to have their concerns and demands presented substantively. Less space is given to protesters' quotes; more space to official sources. Although my work captures amazing individual pieces of journalism that explore themes such as civil rights, protesters' motivations and communities' grief, the dominant narrative accentuates trivial, disruptive and combative actions. My early analyses hint that practices improved during the wake-up call that was 2020, but not by much.



## [What the data say about police brutality and racial bias — and which reforms might work](#)

In 2017, more than half the coverage of immigration, health and science demonstrations included protesters' grievances. Less than one-quarter of Black civil-rights protest coverage did so. After a police officer shot Michael Brown in 2014 in Missouri, one-third of articles emphasized disruption and confrontation. Fewer than 10% described protesters' demands for reform, and then did so shallowly. Our sample found broad consistency across the *St. Louis Post-Dispatch* and top newspapers such as *The New York Times*, *The Wall Street Journal*, *USA TODAY* and *The Washington Post*. The pattern persists over national and local papers and broadcast coverage, as well as around the world.

Activists' work here in Minneapolis, where Floyd died, extends well before and after the events of May 2020, and is often done by people carrying the trauma of their own losses. Many are veterans of demonstrations that followed the deaths of dozens of Black people, including those of two other young Minnesotan men, Philando Castile and Jamar Clark. What I learnt from being on the ground is just how much mainstream media does not cover. Namely, how organized, civil, inspiring and restorative many protest efforts are — from setting up food drives to holding public vigils.

In our preliminary analysis of cable news and Associated Press coverage from May to December 2020, there's a small rise (12% of coverage) in

mentions of police violence during protests from years past. Otherwise, there is little change. Headlines such as “Police violence is just the tip of the issue”, and “Lawmakers use protest momentum to push state racial reforms”, made up only about 69 of 690 articles. Headlines focusing on protester violence and disruption were about four times more common. There were days when some protesters were violent or used radical tactics, but there were solid weeks of peaceful demonstrations. Descriptions of the latter appear in only 4.9% of articles.



### [Wanted: better data on police shootings to reduce mistrust between the police and the communities they serve](#)

Consistently under-represented from the eight years' worth of coverage my team has worked on — from newspapers, websites and TV — are discussions about how racism intersects with other issues. For example, the connection between police shooting Black people and gun violence is rarely made. Stories about police violence against Black and trans women are often pushed to the margins.

Before 2020, journalists' reaction to my research was usually indifference. As newsrooms around the country made efforts to reckon with their racist pasts, they were more willing to engage in initiatives, training courses and workshops. This shift makes reanalysis essential.

My collaborators and I hope to expand our methods to develop more context-specific computer-learning techniques that might be able to detect the nuances in language, such as political dog whistles and coded wording, that especially marginalize Black communities and activist efforts. This would allow us to process information more quickly, perhaps in real time, and to draw broader conclusions. We also hope to broaden analysis of other information sources, including new anti-racism-oriented publications such as *The Emancipator*, or social-media conversations. We want to understand the tensions between narratives from mainstream media and from ‘citizen journalism’ produced by activists and advocates.

Journalists can and do critique individual articles, but their newsrooms often lack the resources to analyse their work as a corpus. Researchers can help them to improve, and hold them to account.

Nature **593**, 315 (2021)

doi: <https://doi.org/10.1038/d41586-021-01314-2>

## Jobs from Nature Careers

- - - [All jobs](#)
    - 
    - [Post-Doc - Numerical and microfabrication development of 4D metamaterials for mechanical, acoustic and photonics applications](#)

[Université Bourgogne Franche-Comté \(UBFC\)](#)

[BESANCON, France](#)

[JOB POST](#)

- [Postdoctoral Position - Ecological Modeler](#)

The University of British Columbia (UBC)

Vancouver, Canada

JOB POST

- **Postdoctoral Fellow | Zandstra Stem Cell Bioengineering Lab**

The University of British Columbia (UBC)

Vancouver, Canada

JOB POST

- **Personalreferent\*in (d/m/w), Gruppe Personaladministration**

Helmholtz Centre for Heavy Ion Research GmbH (GSI)

Darmstadt, Germany

JOB POST

Download PDF

Download PDF

## Related Articles



- **What the data say about police brutality and racial bias — and which reforms might work**



- [Wanted: better data on police shootings to reduce mistrust between the police and the communities they serve](#)

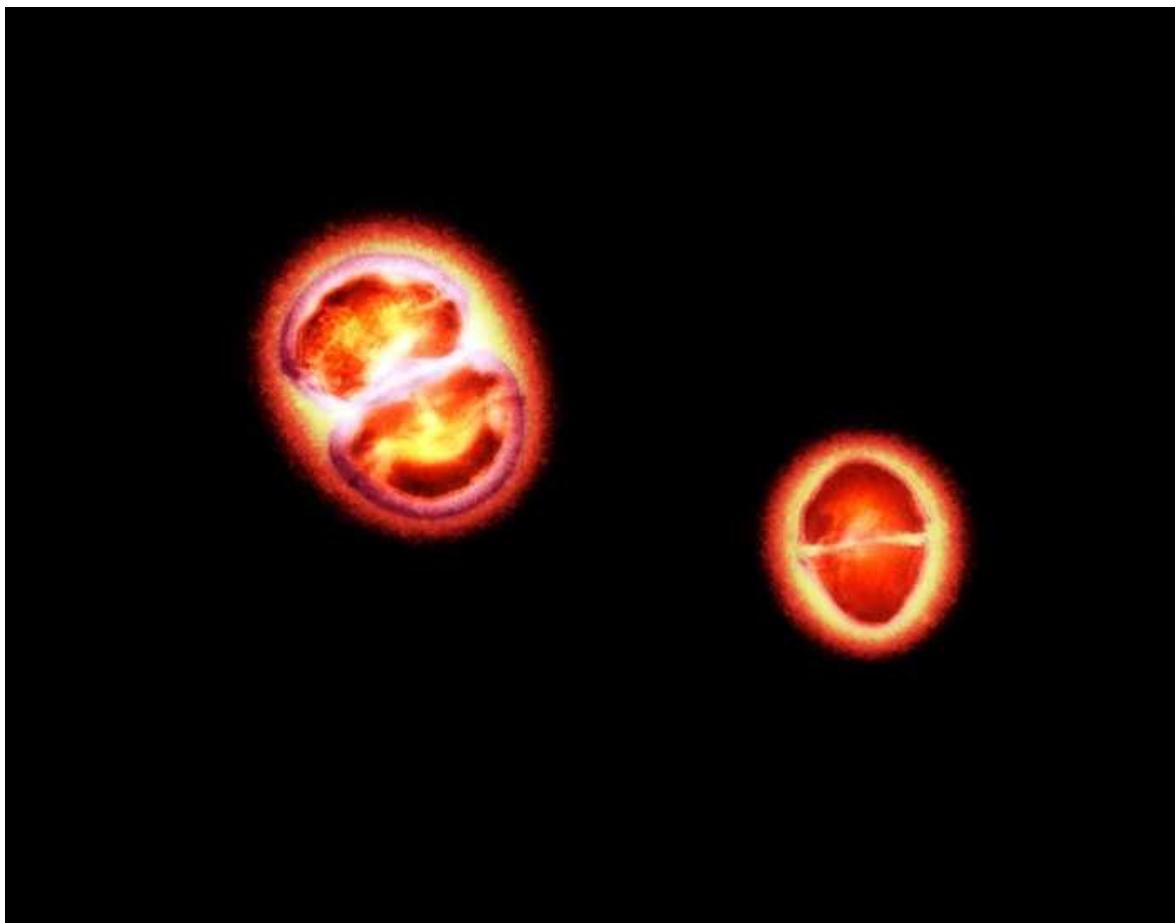


- [Grieving and frustrated: Black scientists call out racism in the wake of police killings](#)

---

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-021-01314-2>

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



Genomic analysis identified starch-loving *Streptococcus sanguinis* bacteria (artificially coloured) in the mouths of modern humans and Neanderthals, but not in chimpanzees' mouths. Credit: National Infection Service/Science Photo Library

Microbiome

14 May 2021

## Microbes in Neanderthals' mouths reveal their carb-laden diet

Gunk on ancient teeth yields bacterial DNA, allowing scientists to trace the oral microbiome's evolution.

Neanderthals' mouths teemed with bacteria that break down starchy food, suggesting that a carbohydrate-rich diet has ancient roots in the human family tree.

An animal's diet, genetics and habits can all shape the diverse set of microscopic organisms in its mouth. To probe the history of humans' oral microbiome, James Fellows Yates at the Max Planck Institute for the Science of Human History in Jena, Germany, and his colleagues sequenced the genomes of microbes scraped from the teeth of primates and hominids, lineages of the tree of life separated by 40 million years. This period encompasses humans' shift to eating farmed plants, and other changes that could have altered microbial diversity.

The team found the same ten types of bacterium in modern humans, Neanderthals, monkeys and apes, pointing to the animals' common origin. But Neanderthals and modern humans — both members of the genus *Homo* — harboured bacteria that the others did not, including a group of *Streptococcus* bacteria, which often help to digest starches.

The genes that enable these *Streptococcus* bacteria to convert starches into energy-rich sugars were much more abundant in modern humans than in Neanderthals, hinting that reliance on starches grew during the course of human evolution.

[Proc. Natl Acad. Sci. USA \(2021\)](#)

- [Microbiome](#)

---

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-021-01260-z>



Proper form: a particular brain region encodes the correct usage of implements such as knives. Credit: Getty

Neuroscience

13 May 2021

## Holding a tool wrong? This brain region will notice

The part of the brain that assesses tool usage is normally more focused on hands than on tools.

Brain regions that play a part in visualizing hands also respond to correct usage of tools, such as the grasping of a spoon by its handle.

Previous studies have shown that different brain regions become active when people look at photographs of hands compared with when they view images of tools. But Stéphanie Rossit at the University of East Anglia in Norwich, UK, and her colleagues found that when it comes to the correct grasping of tools, the visual region for hands is more important than the visual region for tools.

The team scanned the brains of 20 people as they manipulated a spoon, a knife and a pizza-cutter. The participants were also given elongated shapes that did not represent tools, and were asked to grasp the objects by their handles.

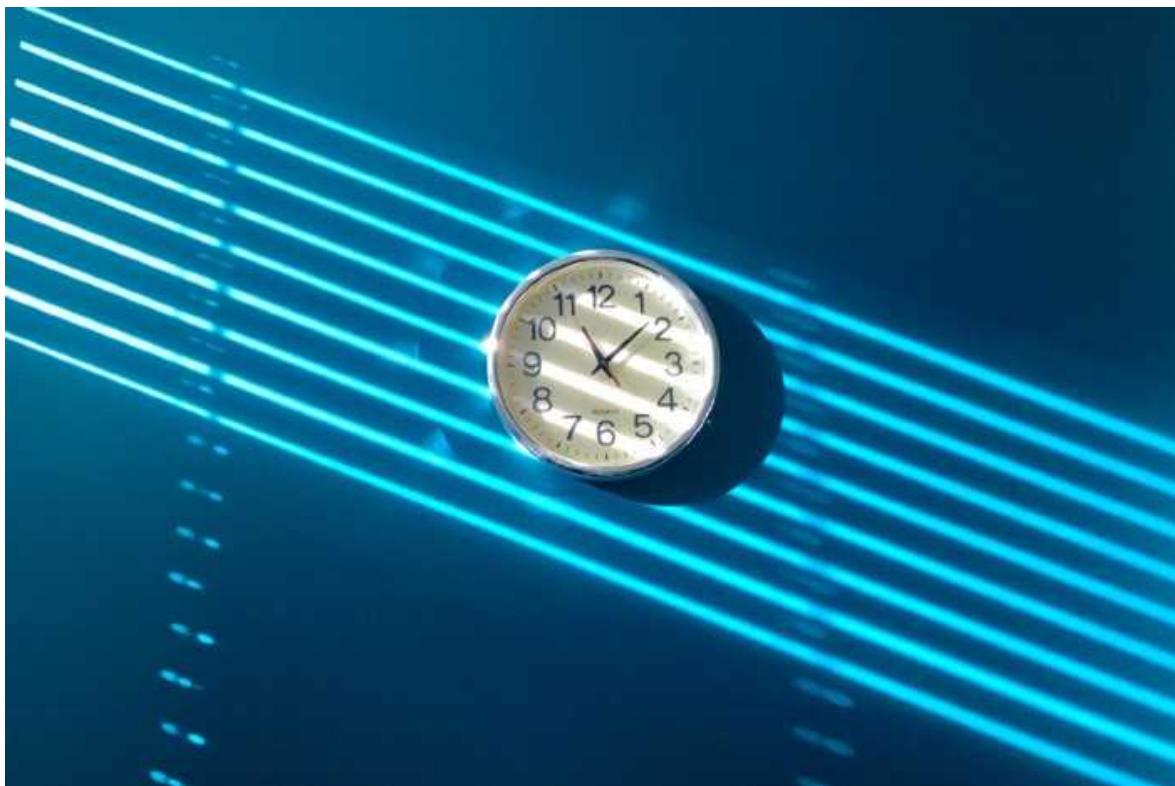
The researchers found that brain regions specialized for visualizing hands could ascertain whether or not people grasped a tool correctly. This was not true when the participants grasped the non-tool objects. The findings provide a better understanding of how the brain evolved to support the use of tools — a defining feature of the human species.

[J. Neurosci. \(2021\)](#)

- [Neuroscience](#)

---

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-021-01261-y>.



The fleeting passage of time is most accurately measured by a clock that emits large amounts of heat. Credit: Getty

Condensed-matter physics

13 May 2021

## How to tell the thermodynamic cost of time

Quantum and ordinary clocks alike gain accuracy as they give off more heat.

An experiment reveals that a clock's accuracy increases as the amount of heat it gives off rises.

Like all machines, clocks consume energy and produce heat. Quantum clocks, which typically keep time by counting the vibrations of single atoms, have an accuracy that depends on how much heat they emit. But do ordinary clocks share this property?

To address this question, Natalia Ares at the University of Oxford, UK, and her colleagues built a simple clock consisting of an ultrathin membrane integrated into an electronic circuit. They powered the clock simply by heating the membrane, making it vibrate. Each oscillation of the membrane represented one tick.

The researchers then tracked the energy going into and out of the clock while varying its operating conditions. They found that the device's accuracy was directly proportional to the amount of heat it released.

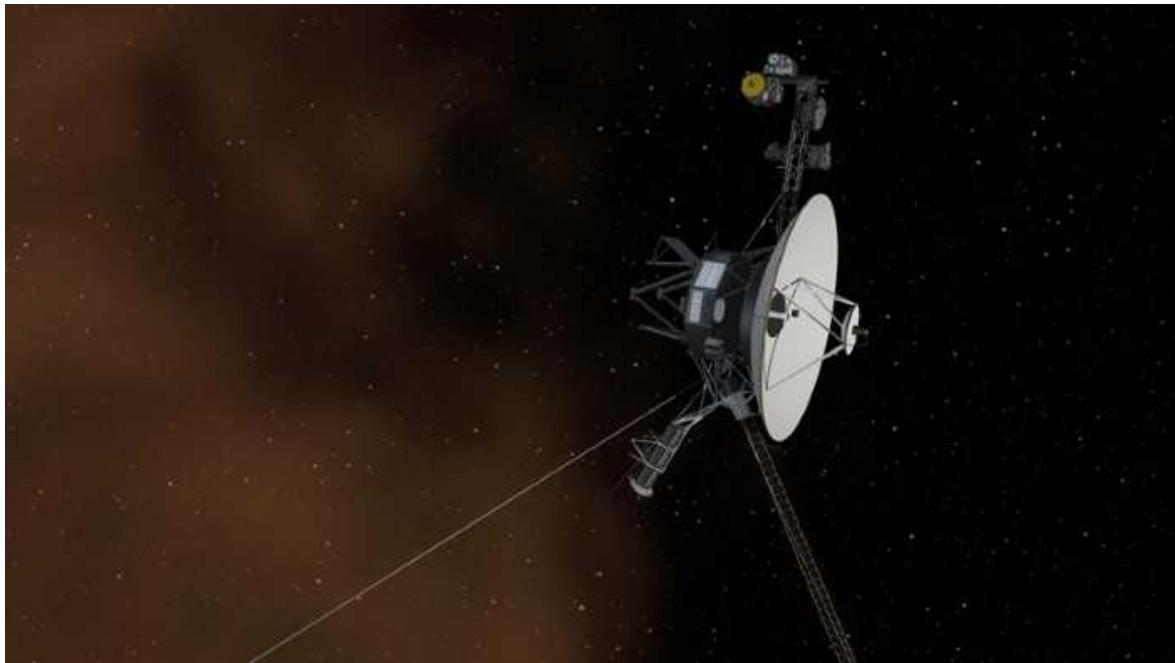
This discovery indicates that the laws of thermodynamics, which place a fundamental limit on the performance of conventional heat engines, also set an intrinsic limit on the accuracy of both quantum and ordinary clocks.

*[Phys. Rev. X \(2021\)](#)*

- [Condensed-matter physics](#)

---

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-021-01259-6>



Data collected by the Voyager 1 spacecraft, which launched in 1977, has helped scientists to calculate the density of the interstellar plasma. Credit: NASA/JPL-Caltech

Astronomy and astrophysics

13 May 2021

## Voyager 1 captures faint ripples in the stuff between the stars

The first spacecraft to visit interstellar space has now become the first to make continuous measurements of waves in that remote realm.

The Voyager 1 spacecraft has detected persistent ripples in the interstellar plasma, through which it has been travelling since it left the Solar System in 2012. By measuring these waves, astrophysicists have made the first continuous measurements of the density of the interstellar plasma, the rarefied medium between the stars.

Launched by NASA 44 years ago, Voyager 1 became the first human-made object to enter interstellar space — the region between star systems — and is still motoring along. Stella Ocker at Cornell University in Ithaca, New York, and her collaborators detected the interstellar-plasma waves by examining regular variations that Voyager recorded in the electric field it encounters as it flies away from the Solar System.

The waves consist of displacements between the plasma's two components: positively charged ions and negatively charged electrons. Despite this displacement, the plasma tends to stay put and the waves go in no particular direction — like the stationary ripples in a lake on a windy day, Ocker says.

In the past, astrophysicists had made similar measurements of interstellar-plasma waves that had been triggered by solar events, but this is the first time that they have measured plasma density continuously.

*Nat. Astro. (2021)*

- [Astronomy and astrophysics](#)

---

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-021-01258-7>



No reason to feel blue: plants with azure blooms (such as *Gentiana ligustica*, above) get preferential treatment from researchers. Credit: M. Adamo *et al./Nature Plants*

Plant sciences

11 May 2021

## Flashy plants draw outsize share of scientists' attention

Blue-flowered plants get the most scientific love; those with green or brown flowers, not so much.

Balconies and gardens are not the only places full of pretty flowers. An analysis of dozens of scholarly articles suggests that visually appealing plants are featured more often in scientific studies than are their nondescript relatives.

Martino Adamo at the University of Torino in Italy and his colleagues analysed 280 studies, published between 1975 and 2020, that focused on plant species typical of the southwestern Alps. The researchers found that eye-catching plants, rather than rare or endangered ones, tended to attract scientists' attention.

Plants with blue flowers were the most studied, and plants with white, red or pink blossoms were investigated more often than were those with brown or green flowers. Scientists also tended to examine plants with taller stems — probably because their flowering parts are more easily accessible than are those of plants with shorter stems.

This ‘aesthetic bias’ could sway conservation efforts in favour of attractive plants, resulting in a lack of interest in less charming, yet often more endangered, species, the authors say.

*[Nature Plants \(2021\)](#)*

- [Plant sciences](#)

---

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-021-01263-w>



Researchers used stem cells (artificially coloured) to ferry remedial genes into the bodies of children with severely compromised immune systems.  
Credit: Science Photo Library

Gene therapy

14 May 2021

## Kids with ‘bubble baby disease’ see gains after gene therapy

Stem cells carrying corrective genes offer hope for a devastating immune disorder.

Children born with the genetic condition known as ‘bubble baby disease’ have no functioning immune system, meaning that even a normally mild infection can be fatal. But in small clinical trials, a form of gene therapy

rebuilt the immune system of such children so well that 96% of them were able to discontinue their main medication for the disease.

More formally known as severe combined immunodeficiency (SCID), the disease drew its nickname from the protective plastic bubbles that were once used to cocoon infants with the condition.

One common form of SCID is caused by a mutation in the gene encoding the enzyme adenosine deaminase (ADA). Current treatments are flawed: enzyme-replacement therapy does not fully restore the immune system, and a previously approved gene therapy caused leukaemia in one recipient.

Donald Kohn at the University of California, Los Angeles, and his colleagues used a virus designed to minimize the risk of cancer-causing effects to deliver a functional *ADA* gene into the participants' own stem cells. Of the 50 children with ADA-SCID who were infused with the cells, 48 were able to stop enzyme-replacement therapy, and none showed signs of cancer.

[N. Engl. J. Med. \(2021\)](#)

- [Gene therapy](#)
- 

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-021-01300-8>



Earthquake damage in southern California. Friction between a quake's moving rocks can raise their temperatures above 1,200 °C. Credit: Etienne Laurent/EPA-EFE/Shutterstock

Geophysics

13 May 2021

## A deep look into the physics of earthquake slip

Laboratory experiments provide a glimpse of what happens when rocks fail.

When two sides of a geological fault grind past one another during an earthquake, the friction can heat the rocks to at least 1,250 °C.

In their laboratory at the National institute of Geophysics and Volcanology in Rome, Stefano Aretusini and colleagues got an unprecedented look at what happens along the frictional zone in which rocks move during a quake. The scientists studied Carrara marble, prized for its use in sculptures and buildings, which belongs to a category of rock that can break along a geological fault.

The team forced two pieces of marble to slide past one another at a speed and level of friction that would be typical during a quake, and measured the temperature of the surfaces every thousandth of a second. The highest temperature recorded by the researchers was 1,250 °C; this exceeds measurements made previously in similar experiments.

Heat within the frictional zone helps to change a rock's physical properties, making it weaken and break during a quake.

*[Geo. Res. Lett. \(2021\)](#)*

- [Geophysics](#)

---

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-021-01262-x>

# News in Focus

- **[Fukushima's wastewater, pandemic response and COVID treatments](#)** [ 19 May 2021]  
News Round-Up • The latest science news, in brief.
- **[Coronavirus variants are spreading in India — what scientists know so far](#)** [ 11 May 2021]  
News • Variants including B.1.617 have been linked to India's surge in infections. Researchers are hurrying to determine how much of a threat they pose.
- **[China has landed its first rover on Mars — here's what happens next](#)** [ 14 May 2021]  
News • The Zhurong landing was the biggest test yet of China's deep-space exploration capabilities. Within days, the rover could start to make geological discoveries.
- **[NASA reboots its role in fighting climate change](#)** [ 07 May 2021]  
News • Space agency aims to breathe new life into its Earth-science programme as US President Joe Biden pushes an ambitious climate agenda.
- **[Pfizer COVID vaccine protects against worrying coronavirus variants](#)** [ 06 May 2021]  
News • Data from Qatar provide strongest evidence yet that COVID-19 vaccines can stop strains thought to pose a threat to immunization efforts.
- **[How many COVID deaths are acceptable in a post-pandemic world?](#)** [ 06 May 2021]  
News • Even after mass vaccinations, some hospitalizations and deaths from the coronavirus are inevitable — but opinions differ on how many is too many for a return to relative normality.
- **[The mysterious microbes that gave rise to complex life](#)** [ 19 May 2021]  
News Feature • As scientists learn more about enigmatic archaea, they're finding clues about the evolution of the complex cells that make up people, plants and more.
- **[How COVID is changing the study of human behaviour](#)** [ 18 May 2021]

News Feature • The pandemic is teaching us key lessons about crisis, communication and misinformation, and is spurring changes in the way scientists study public-health questions.

---

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

---

## NEWS ROUND-UP

19 May 2021

# Fukushima's wastewater, pandemic response and COVID treatments

The latest science news, in brief.



Environmental activists protest against the Fukushima wastewater plan in front of the Japanese embassy in Seoul. Credit: Chung Sung-Jun/Getty

## Scientists back Fukushima wastewater release

Japan's proposal to [discharge more than one million tonnes of contaminated water from the ruined Fukushima Daiichi nuclear power station into the](#)

[ocean](#) has been strongly opposed by neighbours including China and South Korea. But scientists say the risks are likely to be minimal.

Japan revealed the proposal in April, prompting South Korea's foreign minister to express "strong regret and serious concerns". According to China's state news agency Xinhua, Chinese scientists also objected, with Liu Senlin at the China Institute of Atomic Energy in Beijing describing it as "extremely irresponsible".

But other scientists, and the International Atomic Energy Agency, say the radiation in the treated water will be very low, and the release will be gradual. "The facts are not telling me that this is something that we should be very worried about," says Jordi Vives I Batlle, a scientist at the Belgian Nuclear Research Centre in Mol.

The earthquake and tsunami that struck Japan in 2011 led to the failure of cooling systems. In the past decade, 1.25 million tonnes of seawater have been pumped through the units to stop the molten fuel debris from overheating. The water has been treated to remove radioactive material and stored in more than 1,000 steel tanks on site.



A patient who spent nearly a month in the COVID section of an intensive-care unit at Sarasota Memorial Hospital in Florida. Credit: Shannon Stapleton/Reuters/Alamy

## COVID-19 treatment trial will now focus on immunity

A landmark [programme to test potential COVID-19 therapies](#) in dozens of countries is restarting, with a fresh roster of treatments. This time, the treatments aim to temper the immune responses that can worsen severe disease, an approach that has already shown promise in people hospitalized with COVID-19.

The clinical trial, named Solidarity and coordinated by the World Health Organization, will test three existing drugs. The first is infliximab, used to treat autoimmune conditions, including Crohn's disease and rheumatoid arthritis. A second is a cancer drug called imatinib, which researchers hope will target both the coronavirus and inflammation. The third is artesunate, an anti-malaria drug with potential anti-inflammatory effects. All three drugs were chosen because of the promise they showed in smaller clinical trials and their widespread availability, says John-Arne Røttingen, the chair of Solidarity's steering committee.

The trial's original run, which began in March 2020, tested four antiviral treatments (remdesivir, interferon, the malaria drug hydroxychloroquine, and a combination of HIV drugs). None of them was found to save lives or shorten hospital stays.



More than 400,000 people in Brazil have died of COVID-19 since the beginning of the pandemic. Credit: Michael Dantas/AFP via Getty

## **Major investigation reveals How the world failed to curb COVID**

The World Health Organization (WHO) was too cautious in communicating the risks of COVID-19 early last year, according to the [first major investigation of the global pandemic response](#). Had it been bolder, and had nations heeded its guidance, the pandemic might have been curtailed, say the authors of the report.

Last year, at the annual World Health Assembly, countries demanded that the WHO initiate an independent review of how the crisis unfolded, so it could draw lessons for the future. The resulting report, released on 12 May, was assembled by a panel of 13 global-health experts independent of the WHO.

The lengthy investigation identifies February 2020 as the month when the devastating toll of the pandemic might have been lessened, had countries

acted fast. It goes on to list concrete actions that could help to prevent a similar fate should another pathogen with pandemic potential emerge.

Some researchers say that the panel's suggestions for how to strengthen the WHO are too vague. But the panel does make a few ambitious recommendations, including creating a council of world leaders dedicated to fighting pandemics, says Stephen Morrison, director of global health policy at the Center for Strategic and International Studies in Washington DC.

A handful of Asian countries made moves in February last year to curb COVID-19, including instituting testing for SARS-CoV-2. "But the rest of the world sat on their hands," says panel member Joanne Liu, a health-emergency specialist at McGill University in Montreal, Canada. She and her colleagues assessed how the WHO communicated risk, and decided that its cautious weighing of incomplete evidence might help to account for why countries failed to take action.

"When it became obvious that the countries that were wearing masks were faring better than the ones that weren't," she says, "the WHO might have said that even though we don't have all of the data, we should apply the precautionary principle," and recommend masks.

WHO director-general Tedros Adhanom Ghebreyesus announced that he would review the investigation's critiques and proposals, and would discuss reforms with all countries comprising the WHO.

Nature **593**, 319 (2021)

doi: <https://doi.org/10.1038/d41586-021-01315-1>

## **Jobs from Nature Careers**

- - - [All jobs](#)
  - -

**Post-Doc - Numerical and microfabrication development of 4D metamaterials for mechanical, acoustic and photonics applications**

Université Bourgogne Franche-Comté (UBFC)

BESANCON, France

JOB POST

**▪ Postdoctoral Position - Ecological Modeler**

The University of British Columbia (UBC)

Vancouver, Canada

JOB POST

**▪ Postdoctoral Fellow | Zandstra Stem Cell Bioengineering Lab**

The University of British Columbia (UBC)

Vancouver, Canada

JOB POST

**▪ Personalreferent\*in (d/m/w), Gruppe Personaladministration**

Helmholtz Centre for Heavy Ion Research GmbH (GSI)

Darmstadt, Germany

JOB POST

Download PDF

## [Download PDF](#)

---

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-021-01315-1>

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

NEWS

11 May 2021

# Coronavirus variants are spreading in India — what scientists know so far

Variants including B.1.617 have been linked to India's surge in infections. Researchers are hurrying to determine how much of a threat they pose.

- [Gayathri Vaidyanathan](#)
  1. Gayathri Vaidyanathan  
[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)



People queue to refill oxygen cylinders for people with COVID-19 at overwhelmed medical facilities in New Delhi. Credit: Ishant Chauhan/AP/Shutterstock

Scientists are working to understand several coronavirus variants now circulating in India, where a ferocious second wave of COVID-19 has devastated the nation and caught authorities unawares. The country recorded nearly 400,000 new infections on 9 May, taking its total to more than 22 million (see ‘Surging cases of COVID-19’).

Evidence is growing that one variant first detected in India might be more transmissible and slightly better at evading immunity than existing variants. Animal models also hint that it might be able to cause more severe disease. Researchers want to know if this variant and others might be driving the second wave and what kind of danger they pose globally.

In just a few weeks, the B.1.617 variant has become the dominant strain across India and has spread to about 40 nations, including the United

Kingdom, Fiji and Singapore.

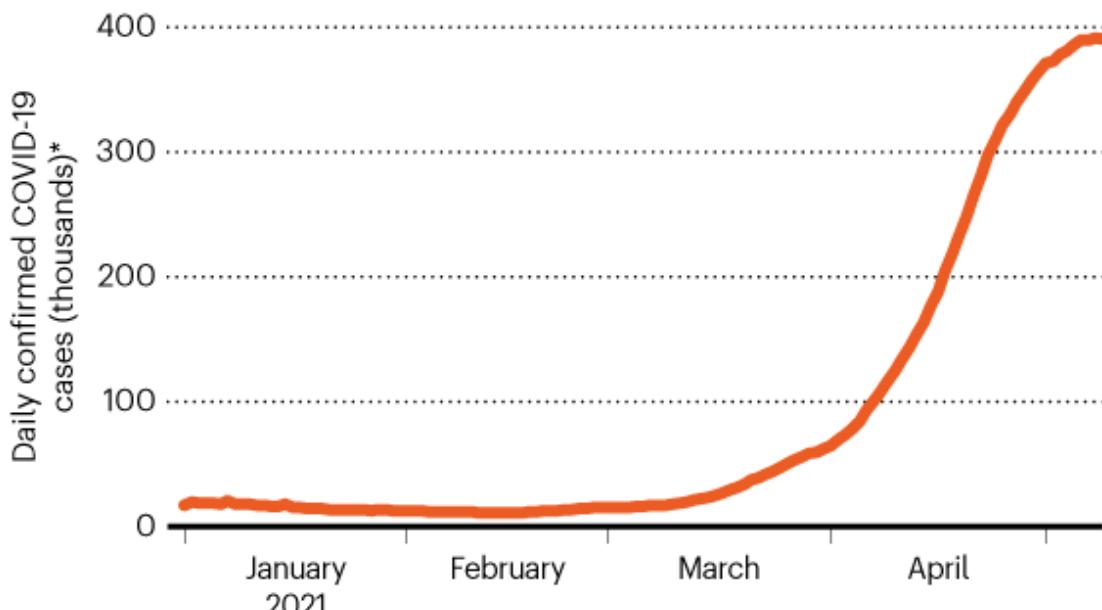
## A growing problem

Two weeks ago, it looked as if multiple variants were behind a series of surges in India. Genomic data indicated that B.1.1.7, first identified in the United Kingdom, was dominant in Delhi and the state of Punjab, and a new variant dubbed B.1.618 was present in West Bengal. B.1.617 was dominant in Maharashtra.

But since then, B.1.617 has overtaken B.1.618 in West Bengal, has become the leading variant in many states, and is increasing rapidly in Delhi. “In some states, the surge can be tied to 617,” Sujeev Singh, director of the National Centre for Disease Control, based in New Delhi, told journalists on 5 May.

### SURGING CASES OF COVID-19

New daily cases in India have been rapidly on the rise since March and have now far surpassed a peak of around 100,000 new cases per day last September.



Source: Our World in Data

Some say this could indicate that the variant is highly transmissible. “Its prevalence has increased over other variants in much of India, suggesting that it has better ‘fitness’ over those variants,” says Shahid Jameel, a virologist at Ashoka University in Sonipat who chairs the scientific advisory group of the Indian SARS-CoV-2 Genome Sequencing Consortia (INSACOG).

Ravindra Gupta, a virologist at the University of Cambridge, UK, agrees that it is “highly likely to be more transmissible”.



### [India's massive COVID surge puzzles scientists](#)

On Monday, the World Health Organization (WHO) designated B.1.617 a ‘variant of concern’. Variants are classified in this way when there is evidence that they spread more rapidly, cause more severe disease or evade previously acquired immunity better than do circulating versions of the virus. On 7 May, the UK government declared the B.1.617.2 subtype a variant of concern in the United Kingdom. It revealed that recorded B.1.617.2 infections in the country had risen from 202 to 520 in a single week.

Several other variants of concern have had a significant impact globally. These include B.1.351, which was identified in South Africa in late 2020; studies suggesting that the University of Oxford–AstraZeneca jab is less effective against that variant led to the nation suspending its roll-out. Similarly, the P.1 variant, which is able to evade some immunity, contributed to a major second wave in Brazil early this year. And the highly transmissible B.1.1.7 strain emerged in the United Kingdom in late 2020 and led to a surge of cases there and elsewhere.

## Emerging mosaic

Data on B.1.617 are only just trickling out, but a mosaic of findings hints that it has an edge over variants already circulating in India.

Indian scientists first detected B.1.617 in a few samples in October. INSACOG ramped up surveillance in late January in response to a rising number of variants, and scientists noticed that B.1.617 was on the rise in Maharashtra. By mid-February, it accounted for 60% of cases there, says Priya Abraham, director of the National Institute of Virology (NIV) in Pune. Since then, multiple sub-lineages have emerged.

In a detailed genomic and structural analysis of B.1.617 published as a preprint<sup>1</sup> on 3 May, NIV scientists identified eight mutations in the virus's spike protein, through which it gains entry to cells. Two of them look similar to mutations that have allowed other variants of concern to become more transmissible, and a third resembles a mutation that might have allowed P.1 to partially evade immunity.



People with COVID-19 recuperate at a care centre in Delhi. Credit: Raj K. Raj/*Hindustan Times*/Shutterstock

The genomics work was backed up days later by a preprint<sup>2</sup> from a team in Germany, which shows that B.1.617 is moderately better than an earlier variant at entering human intestine and lung cells in the lab.

It is unclear whether this “minor” advantage could lead to more spread in the real world, says lead author Markus Hoffman, an infection biologist at the Leibniz Institute for Primate Research in Göttingen.

Small studies in animals suggest that the variant could cause more severe disease. In a 5 May preprint<sup>3</sup>, a team led by virologist Pragya Yadav at the NIV found that hamsters infected with B.1.617 had more inflammation in their lungs than did animals infected with other variants.

## Disease-causing potential

Gupta says this research shows that B.1.617 has enhanced potential to cause disease. But he cautions that “it’s difficult to extrapolate from hamsters to humans”, and says that data on disease severity in people are needed.

Research<sup>4</sup> from Gupta’s own lab suggests that antibodies are slightly less effective against the variant than against others. The team collected blood serum from nine people who had received one dose of the Pfizer vaccine and tested it against a harmless carrier virus modified to contain the SARS-CoV-2 spike protein, with the mutations from B.1.617. Serum from vaccinated individuals typically contains antibodies that can block, or ‘neutralize’, the virus and prevent cells from getting infected.



### India's COVID-vaccine woes — by the numbers

Gupta’s team discovered that neutralizing antibodies generated by the vaccinated individuals were about 80% less potent against some of the mutations in B.1.617, although this would not render vaccination ineffective, he says. The researchers also found that some health-care workers in Delhi who had been vaccinated with Covishield, an Indian version of the Oxford–AstraZeneca vaccine, had become reinfected, with most cases tied to B.1.617.

Similarly, the German team tested<sup>2</sup> serum from 15 people who had previously been infected with SARS-CoV-2, and found that their antibodies

neutralized B.1.617 about 50% less effectively than previously circulating strains. When they tested serum from participants who'd had two shots of the Pfizer vaccine, they found that the antibodies were about 67% less potent against B.1.617.

Two other small studies, one from Yadav's team<sup>5</sup> testing the Covaxin vaccine made by Indian firm Bharat Biotech in Hyderabad, and an [as-yet-unpublished study](#) on Covishield, showed that the vaccines continue to work. But Yadav observed small drops in the effectiveness of neutralizing antibodies generated by the Covaxin vaccine.

The B.1.617 variant does seem to have an advantage over previously circulating versions of the virus, especially in individuals whose immunity is waning a while after previous infection or vaccination, says Hoffman.

## Caveats and caution

But Gupta cautions that these lab studies all involve small groups and show smaller drops in antibody effectiveness, compared with what has been seen with other variants of concern.



[India will supply coronavirus vaccines to the world — will its people benefit?](#)

Scientists also warn that experiments in serum are not always a good guide to whether a variant can evade immunity from a vaccine in the real world. Vaccines can cause the production of vast quantities of antibodies, so a dip in potency might not be significant. Furthermore, other parts of the immune system, such as T cells, might not be affected.

For example, the B.1.351 variant has been linked to much steeper drops in the potency of neutralizing antibodies, but studies in humans suggest that many vaccines remain highly effective against that variant, particularly at preventing severe disease.

For these reasons, the vaccines are likely to remain effective against B.1.617 and to limit severe disease. “The vaccine is still working,” says Yadav. “If you get vaccinated, you “will be protected, and the severity will be less”.

Nevertheless, “the surge in cases in India and scenes witnessed there is of grave concern internationally”, Nick Loman, a microbial genomicist and bioinformatician at the University of Birmingham, UK, told the Science Media Centre in London after the United Kingdom declared B.1.617.2 a variant of concern. “This variant will now be one to watch carefully.”

Nature **593**, 321-322 (2021)

doi: <https://doi.org/10.1038/d41586-021-01274-7>

## References

1. 1.

Cherian, S. *et al.* Preprint at bioRxiv  
<https://doi.org/10.1101/2021.04.22.440932> (2021).

2. 2.

Hoffmann, M. *et al.* Preprint at bioRxiv  
<https://doi.org/10.1101/2021.05.04.442663> (2021).

3. 3.

Yadav, P. D. *et al.* Preprint at bioRxiv  
<https://doi.org/10.1101/2021.05.05.442760> (2021).

4. 4.

Ferreira, I. *et al.* Preprint at bioRxiv  
<https://doi.org/10.1101/2021.05.08.443253> (2021).

5. 5.

Yadav, P. D. *et al.* *Clin. Infect. Dis.* <https://doi.org/10.1093/cid/ciab411> (2021).

## **Jobs from Nature Careers**

- - - [All jobs](#)
    - - [Post-Doc - Numerical and microfabrication development of 4D metamaterials for mechanical, acoustic and photonics applications](#)  
[Université Bourgogne Franche-Comté \(UBFC\)](#)  
[BESANCON, France](#)  
[JOB POST](#)
      - [Postdoctoral Position - Ecological Modeler](#)  
[The University of British Columbia \(UBC\)](#)  
[Vancouver, Canada](#)  
[JOB POST](#)

- [Postdoctoral Fellow | Zandstra Stem Cell Bioengineering Lab](#)

[The University of British Columbia \(UBC\)](#)

[Vancouver, Canada](#)

[JOB POST](#)

- [Personalreferent\\*in \(d/m/w\), Gruppe Personaladministration](#)

[Helmholtz Centre for Heavy Ion Research GmbH \(GSI\)](#)

[Darmstadt, Germany](#)

[JOB POST](#)

[Download PDF](#)

[Download PDF](#)

## Related Articles



- [India's massive COVID surge puzzles scientists](#)



- [India's COVID-vaccine woes — by the numbers](#)



- **Why India must tackle a mutating virus head-on**

---

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-021-01274-7>

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

NEWS

14 May 2021

- Update [15 May 2021](#)

# China has landed its first rover on Mars — here's what happens next

The Zhurong landing was the biggest test yet of China's deep-space exploration capabilities. Within days, the rover could start to make geological discoveries.

- [Smriti Mallapaty](#)
  1. Smriti Mallapaty  
[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)



The surface of Mars, photographed by China's Tianwen-1 probe after it arrived in orbit in February.Credit: CNSA/Xinhua/Alamy

China's Tianwen-1 spacecraft, in orbit around the red planet, has dropped its lander and rover — named Zhurong after a Chinese god of fire — completing the most perilous stage of its ten-month mission.

According to Chinese state news agency Xinhua, an entry capsule enclosing the vehicles separated from the orbiter at about 4 a.m. Beijing time on 15 May. After several hours, it entered Mars's atmosphere at an altitude of 125 kilometres.

It then hurtled towards the surface at 4.8 kilometres per second, protected by a heat shield. As the probe closed in on Mars, it released a huge parachute to slow its progress, and then used rocket boosters to brake. Once it reached 100 metres above the Martian surface, it hovered and used a laser-guided system to assess the area for obstacles such as boulders before landing.

The craft's plummet through the Martian atmosphere had to be performed autonomously. "Each step had only one chance, and the actions were closely linked. If there had been any flaw, the landing would have failed," Geng Yan, an official at the Lunar Exploration and Space Program Center of the China National Space Administration (CNSA), told Xinhua.

## 'Big leap for China'

This is China's first mission to Mars, and makes the country only the third nation — after Russia and the United States — to have landed a spacecraft on the planet. The mission "is a big leap for China because they are doing in a single go what NASA took decades to do", says Roberto Orosei, a planetary scientist at the Institute of Radioastronomy of Bologna in Italy.

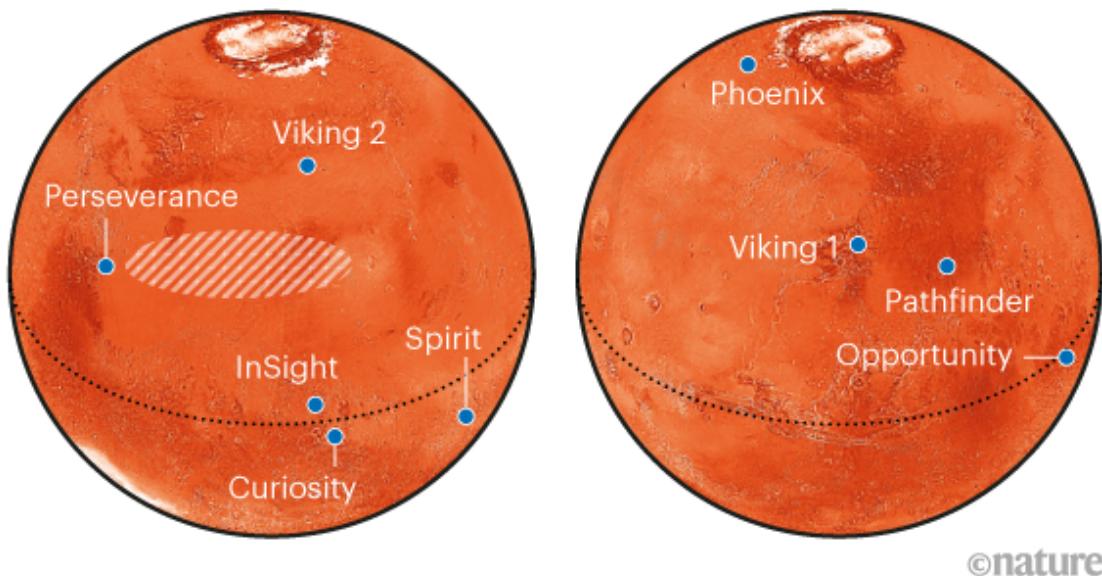
Zhurong now joins several other active Mars missions. NASA's Perseverance rover, which [arrived on 18 February](#), is several hundred kilometres away from the landing site, and NASA's Curiosity rover has been poking around the planet since 2012. Several spacecraft are also circling Mars, including the United Arab Emirates' [Hope orbiter](#), which also arrived in February. "The more the merrier on Mars," says David Flannery, an astrobiologist at Queensland University of Technology in Brisbane, Australia.

Researchers say that the engineering feat of getting there has taken precedence over science in China's first tour of Mars, but the mission could still reveal new geological information. They are especially excited about the possibility of detecting permafrost in Utopia Planitia, the region in the northern hemisphere of Mars where Zhurong has landed (see 'Landing site').

## LANDING SITE

A Chinese rover called Zhurong is set to land in Utopia Planitia in Mars's northern hemisphere. Soviet and US missions have landed in many regions of Mars.

● **Tianwen-1** landing site in Utopia Planitia ● Previous missions



©nature

## Biggest test yet

The Tianwen-1 mission included an orbiter, a lander and a rover — making it the first to send all three elements to the planet. The spacecraft departed Earth in July 2020 and arrived at Mars in February 2021, but the landing was the biggest test yet of China's nascent deep-space exploration capabilities.

Landing on Mars is notoriously difficult, not least because engineers back on Earth have no control over it in real time, and must leave pre-programmed instructions to play out. Many missions have been lost, or have crashed on arrival.

In 1997, NASA's Mars Pathfinder sent its first rover, named Sojourner, to a rocky region of the planet. "We didn't get a lot of amazing science from that mission, but it paved the way for much more capable autonomous rovers,

and now we are reaping the benefits of those missions,” says Flannery, who works on Perseverance, [NASA’s fifth Mars rover](#).

## What to expect

Within days, the six-wheeled solar-powered rover will trundle off the lander to explore for at least three months — but it could survive for years, as NASA’s [Spirit](#) and [Opportunity](#) rovers did.

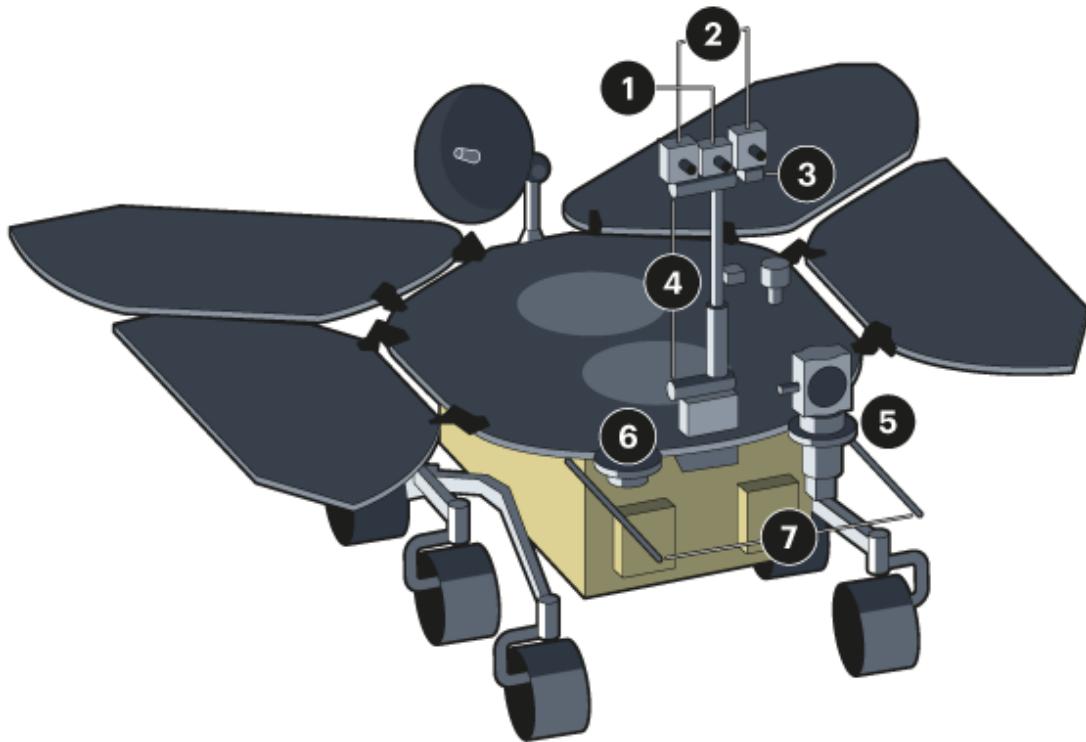
Utopia Planitia, where Zhurong now sits, is a wide, flat expanse in a vast, feature-less basin that formed when a smaller object smashed into Mars billions of years ago.

The basin’s surface is mostly covered in volcanic material, which could have been modified by more-recent processes, such as the repeated freezing and thawing of ice. Orosei says that studies of the region from Mars’s orbit suggest that a layer of permafrost could be hiding just below the surface.

In 1976, NASA’s Viking 2 mission also landed on Utopia Planitia, but farther north of where Zhurong has touched down. “It’s a good place to try a first landing,” Flannery said before the landing. The low altitude, clear terrain and potential for finding ice in the subsurface also means that future missions might be able to collect samples there, and that the region could make a good landing site for crewed missions, he says.

## ZHURONG

The first Chinese rover to land on Mars will explore Utopia Planitia, a large, flat basin in Mars's northern hemisphere. Zhurong will survey the geological structures on, and just below, the surface of the red planet.



1. **MSCam:** Multispectral camera
2. **NaTeCam:** Two cameras for imaging and navigating the terrain
3. **MCS climate station:** Wind and sound detectors
4. **MoMAG:** Magnetometer probes for detecting magnetic field
5. **MarSCoDe:** Spectrometer with 'LIBS' laser for measuring rock composition from a distance
6. **MCS climate station:** Temperature and pressure sensors
7. **RoPeR:** Subsurface radar antennae with two frequency channels

©nature

## Measuring Mars

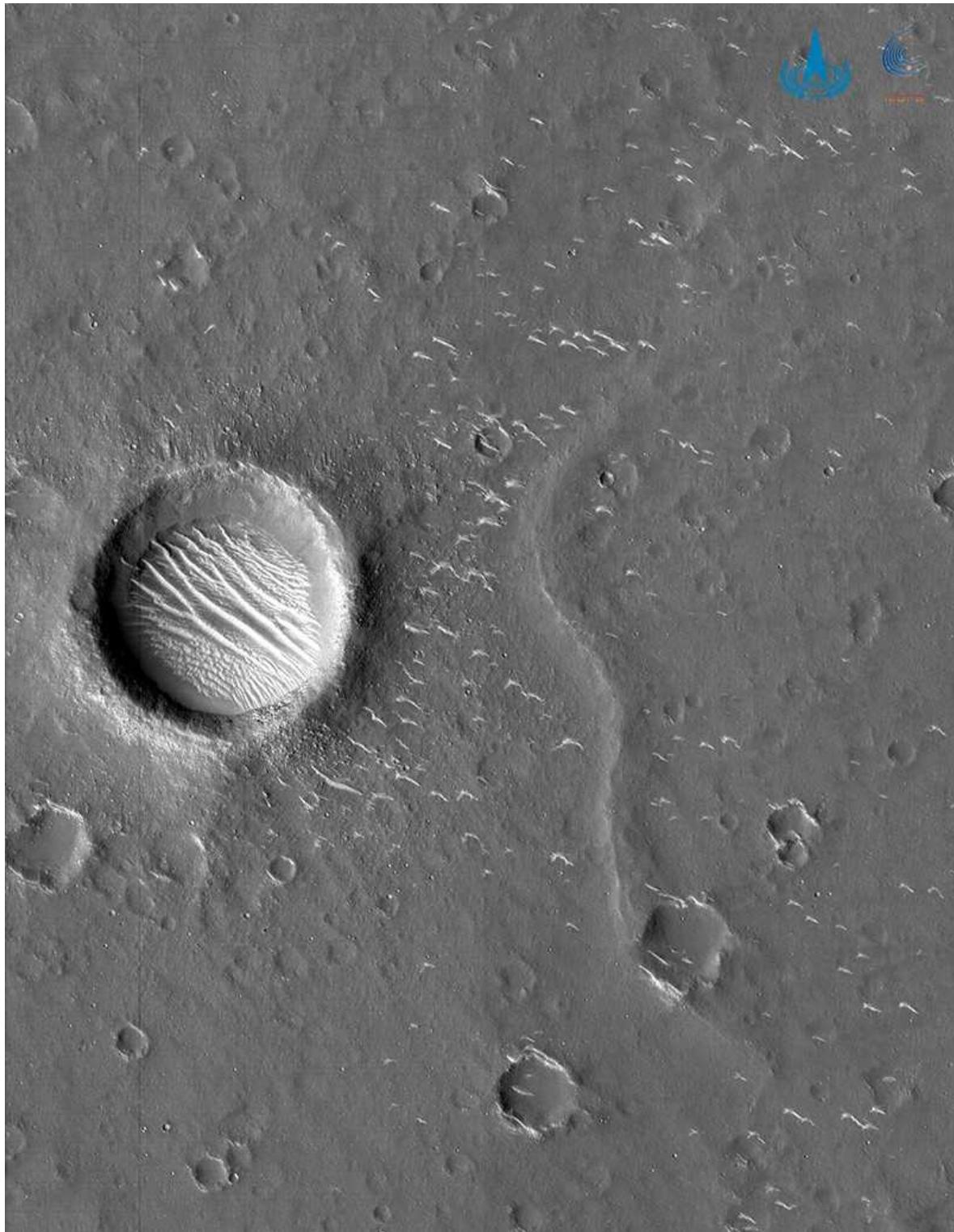
Zhurong is kitted out with a suite of instruments for exploring the Martian environment (see 'Zhurong'). Two cameras are fitted on a mast to take images of nearby rocks while the rover is stationary; these will be used to

plan the journeys that it takes. A multispectral camera placed between these two navigation imagers will reveal the minerals present in these rocks.

Like Perseverance, Zhurong has ground-penetrating radar. As it winds its way across the basin, this will reveal the geological processes that led to the formation of the regions through which the rover travels. With luck, Zhurong might detect the thin horizon that marks any permafrost, says Oro sei. Knowing how deep this lies, and its general characteristics, could offer insights into more recent climate changes on Mars, and reveal the fate of ancient water that could have once soaked the surface, he says.

If the researchers are really fortunate, they might even find some very ancient rocks, which could offer a window into our own planet's history, says Joseph Michalski, a planetary scientist at the University of Hong Kong: most of the similar evidence here on Earth has been destroyed by plate tectonics, says Michalski.

Zhurong's spectrometer includes a laser-based technology that can zap rocks to study their make-up. It will also be the first rover equipped with a magnetometer to measure the magnetic field in its vicinity. The instrument could provide insights into how Mars lost its strong magnetic field, an event that transformed the planet into a cold, dry place, uninviting to life.



A crater on the surface of Mars, captured by China's Tianwen-1 orbiter. Credit: CNSA/Xinhua/Alamy

## Orbital insights

From orbit, [Tianwen-1](#) will communicate Zhurong's insights to Earth. But the orbiter — the name of which means ‘questions to heaven’ — will also make its own scientific contributions with its seven instruments, including cameras, ground-penetrating radar and a spectrometer.

A magnetometer and particle analysers will study the boundary between the higher Martian atmosphere and solar winds to better understand how Mars’s magnetic field operates today. Combined with data from other orbiters studying the planet’s upper atmosphere, this knowledge will offer researchers “a much better picture of what goes on around Mars”, says Orosei.

A successful Mars landing could usher in more-advanced Chinese missions — including a sample-return initiative, which is planned to take place by 2030.

Nature **593**, 323-324 (2021)

doi: <https://doi.org/10.1038/d41586-021-01301-7>

## Updates & Corrections

- **Update 15 May 2021:** This story was updated to reflect the fact that Zhurong has landed successfully on the surface of Mars.

## [Jobs from Nature Careers](#)

- - - [All jobs](#)
    - - [Post-Doc - Numerical and microfabrication development of 4D metamaterials for mechanical,](#)

## **acoustic and photonics applications**

Université Bourgogne Franche-Comté (UBFC)

BESANCON, France

JOB POST

- **Postdoctoral Position - Ecological Modeler**

The University of British Columbia (UBC)

Vancouver, Canada

JOB POST

- **Postdoctoral Fellow | Zandstra Stem Cell Bioengineering Lab**

The University of British Columbia (UBC)

Vancouver, Canada

JOB POST

- **Personalreferent\*in (d/m/w), Gruppe Personaladministration**

Helmholtz Centre for Heavy Ion Research GmbH (GSI)

Darmstadt, Germany

JOB POST

Download PDF

Download PDF

## Related Articles

-  [China's first Mars explorer arrives at the red planet](#)
-  [Countdown to Mars: three daring missions take aim at the red planet](#)
-  [China's successful launch of Mars mission seals global era in deep-space exploration](#)

---

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-021-01301-7>

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

NEWS

07 May 2021

# NASA reboots its role in fighting climate change

Space agency aims to breathe new life into its Earth-science programme as US President Joe Biden pushes an ambitious climate agenda.

- [Alexandra Witze](#)

1. Alexandra Witze

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)



NASA has been involved in Earth-observation studies since 1960, capturing images such as this one from the DSCOVR satellite. Credit: Carl Hostetter/NASA EPIC Team

NASA is best known for exploring other worlds, whether that's [sending astronauts to the Moon](#) or [flying helicopters on Mars](#). But under US President Joe Biden, the space agency intends to boost its reputation as a major player in studying Earth — especially with an eye towards fighting climate change.

“Biden made clear that climate is a priority,” says Waleed Abdalati, director of the Cooperative Institute for Research in Environmental Sciences in Boulder, Colorado. “There’s a clear role for NASA to play in that,” he says, given all the Earth-science research it funds and the Earth-observing satellites it launches.



### [Ageing satellites put crucial sea-ice climate record at risk](#)

In recent months, NASA has signalled its intention to reinvigorate its role in informing US climate policy, by appointing its first climate adviser and ramping up work on key missions to study how Earth’s climate is changing.

The work is particularly crucial as climate change accelerates, agency officials say. “The demand for actionable information is going to increase pretty dramatically over the next decade or two,” says Karen St. Germain, head of NASA’s Earth-science division in Washington DC.

## **Monitoring change**

Among the many US federal agencies that Biden has conscripted to curb climate change, NASA stands out because it is a leader in basic planetary discoveries. Its history of Earth observation stretches back to 1960, when it launched the TIROS-1 satellite to test the feasibility of monitoring weather from space. Over more than six decades, NASA has designed, built and launched spacecraft to [observe Earth as it changes](#). Often working in concert

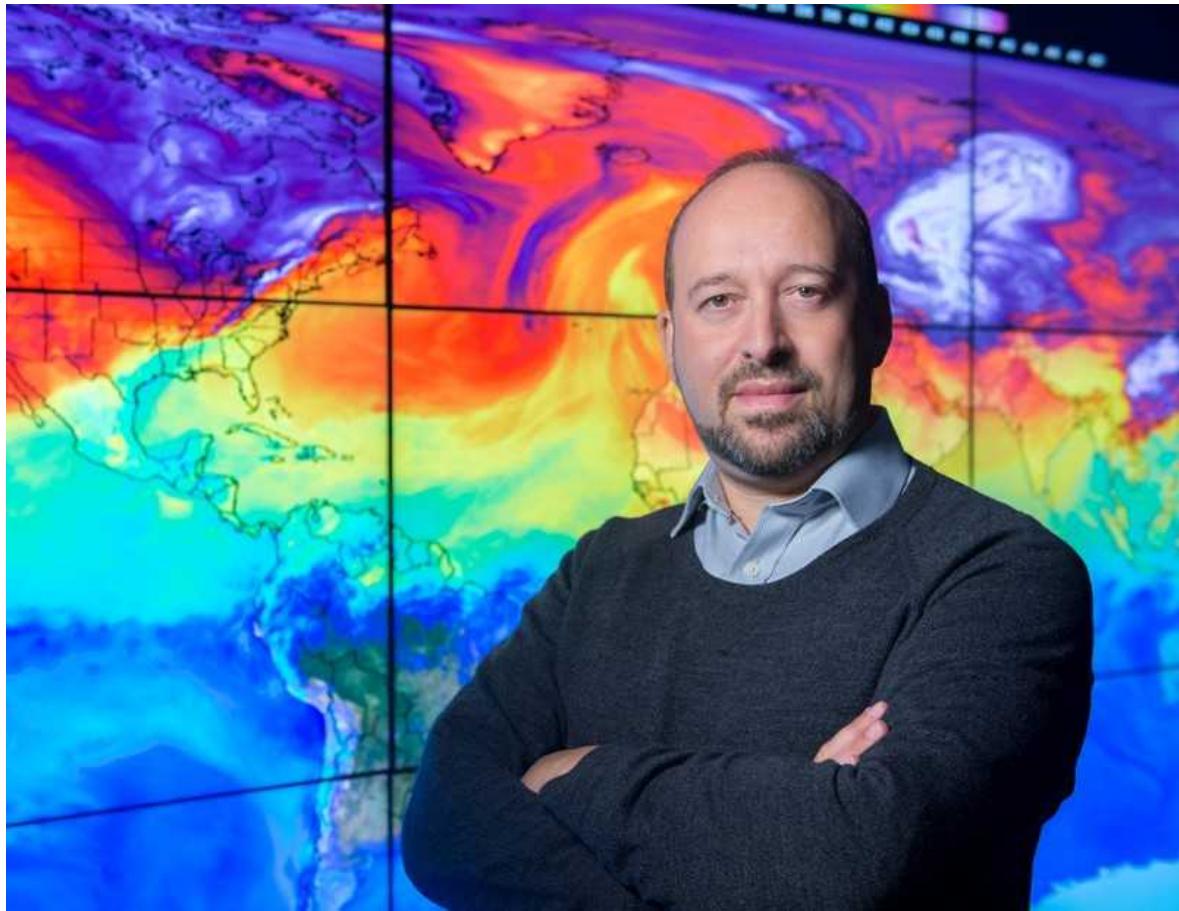
with the US National Oceanic and Atmospheric Administration (NOAA), which has primary responsibility for national weather forecasting, NASA runs satellites that measure ice sheets melting and carbon dioxide flowing through the atmosphere. The agency also flies aeroplanes to gather data about planetary change and funds a broad array of fundamental climate research, such as climate-modelling studies. “Our central role is in understanding how the Earth system is changing,” says St. Germain.



### US urged to invest in sun-dimming studies as climate warms

Biden’s predecessor, Donald Trump, whose [policies favoured industry and downplayed climate change](#), repeatedly tried to cancel major NASA Earth-science missions, only to see them rescued by Congress. It was part of a broader pattern across the Trump administration of undercutting climate-change research and policy. NASA escaped the worst of those attacks by keeping most of its climate-change and Earth-science research below the radar of Trump officials. But it was a politically fraught time for the agency.

Now, NASA is literally reclaiming its seat at the table. Biden initially left the agency off the high-level climate task force he established a week after taking office in January. Following some pointed phone calls, NASA muscled its way into that group, and is now represented alongside administration heavyweights such as the secretaries of the treasury and defence as they discuss the nation’s climate strategy.



Gavin Schmidt is NASA's first climate adviser.Credit: NASA

"If you're going to make policy related to scientific questions, you need to have science at the table," says Gavin Schmidt, a climate modeller at NASA's Goddard Institute for Space Studies in New York City and the agency's new climate adviser.

NASA's new administrator, former senator Bill Nelson, has said that he supports the agency's Earth-science research. "You can't mitigate climate change unless you measure it, and that's NASA's expertise," he said at his Senate confirmation hearing on 21 April.

## Budget struggles

Among NASA's epic observations of climate change are a 29-year programme recording global sea-level rise, measured precisely from space

with French collaborators and others, and studies that began in 2002 to track ice loss from Greenland and Antarctica, done with German partners.

Upcoming missions include a [US-Indian radar satellite](#) that will track planetary changes such as shifts in sea-ice cover, and a US-Canadian-French-UK spacecraft that will survey freshwater resources and ocean currents. Both are slated to launch next year.

But NASA has struggled to get some of its most anticipated climate missions under way. It intends to launch a series of spacecraft that would measure fundamental aspects of global warming, such as shifts in clouds and precipitation, and changes in Earth's mass as groundwater dries up. These missions were shaped by [an influential 2018 report](#) from the US National Academies of Sciences, Engineering, and Medicine, [which named five 'designated observables'](#) that NASA should track. Collectively, NASA calls them an Earth system observatory; they would help scientists to continue to track global change, and give policymakers data they need to inform actions on climate change.



### [Biden pursues giant boost for science spending](#)

The report estimated that the necessary missions might cost between US\$300 million and \$800 million apiece, and suggested that they might be doable even with tight budgets. NASA's annual Earth-science budget has hovered around \$2 billion for years, even as other agency programmes, such as planetary sciences, received huge funding increases.

In April, Biden proposed boosting NASA's Earth-sciences budget to nearly \$2.3 billion, although Congress would need to approve it. "It's a big help and I'm supportive of it," says Abdalati, who previously served as NASA's chief scientist and also co-led the 2018 National Academies report. But "as a result of underinvestment for so many years, it looks better than it is". (NOAA also got a proposed boost of around \$500 million for its own line of weather and climate satellites.)

## Making ends meet

Despite its budgetary struggles, NASA has managed to continue doing climate science in the past few years. It scraped together money to start work on a new instrument called Libera, which will launch in 2027 to measure solar radiation, so that the agency can maintain a crucial four-decade record of how much heat Earth's atmosphere absorbs. But it hasn't made the progress it would have liked on monitoring the designated observables. "Over the last year, we've been in an especially challenging environment with COVID," says St. Germain.



### [NASA probe will track melting polar ice in unprecedented detail](#)

The agency has been trying to save money and speed things up by working with partners such as commercial companies. Sometimes that has worked well. For instance, the agency set up a programme to buy Earth-observation data from satellite companies and make the information available to NASA-funded researchers, to complement measurements from government

satellites. At other times, NASA has been left in the lurch. For example, the agency planned several years ago to pay for a ride on commercial satellite launches for some of its spacecraft that needed to reach geostationary orbit, a relatively high vantage point above Earth. But then the commercial satellite operators changed their plans, leaving a pollution-monitoring spacecraft called TEMPO and a carbon-mapping mission called GeoCarb scrambling for new rides to space.

If Congress approves a significant boost for NASA's Earth-science division, then the agency might finally be able to accelerate progress on satellites to track climate change. The funding decision is expected in the coming months.

"The switch in administration now means that we can actually start to implement some of these missions," says Helen Fricker, a glaciologist at the Scripps Institution of Oceanography in La Jolla, California, who studies Antarctic ice loss. "We can make up for lost time and get on with it."

Nature **593**, 324-325 (2021)

doi: <https://doi.org/10.1038/d41586-021-01230-5>

## Jobs from Nature Careers

- - - [All jobs](#)
    - 
    - [Post-Doc - Numerical and microfabrication development of 4D metamaterials for mechanical, acoustic and photonics applications](#)

[Université Bourgogne Franche-Comté \(UBFC\)](#)

[BESANCON, France](#)

JOB POST

- **Postdoctoral Position - Ecological Modeler**

The University of British Columbia (UBC)

Vancouver, Canada

JOB POST

- **Postdoctoral Fellow | Zandstra Stem Cell Bioengineering Lab**

The University of British Columbia (UBC)

Vancouver, Canada

JOB POST

- **Personalreferent\*in (d/m/w), Gruppe Personaladministration**

Helmholtz Centre for Heavy Ion Research GmbH (GSI)

Darmstadt, Germany

JOB POST

Download PDF

Download PDF

## Related Articles

-  [NASA probe will track melting polar ice in unprecedented detail](#)
-  [Biden pursues giant boost for science spending](#)
-  [Ageing satellites put crucial sea-ice climate record at risk](#)
-  [US urged to invest in sun-dimming studies as climate warms](#)

---

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-021-01230-5>

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

NEWS

06 May 2021

# Pfizer COVID vaccine protects against worrying coronavirus variants

Data from Qatar provide strongest evidence yet that COVID-19 vaccines can stop strains thought to pose a threat to immunization efforts.

- [Ewen Callaway](#)

1. Ewen Callaway

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)



Doses of the Pfizer–BioNTech vaccine being prepared in Toronto, Canada. Credit: Steve Russell/Toronto Star/Getty

Qatar's second wave of COVID-19 was a double whammy. In January, after months of relatively few cases and deaths, the Gulf nation saw a surge driven by the fast-spreading B.1.1.7 variant, which was first identified in the United Kingdom. Weeks later, the B.1.351 strain, which is linked to reinfections and dampened vaccine effectiveness, took hold.

Amid this storm, researchers in Qatar have found some of the strongest evidence yet that current vaccines can quell variants such as B.1.351. Clinical trials in South Africa — where B.1.351 was first identified — had suggested that vaccines would take a hit against such variants. But this study offers a fuller picture of what countries battling such variants can expect.



### Who received the first billion COVID vaccinations?

People in Qatar who received two doses of the Pfizer–BioNTech vaccine were 75% less likely to develop a case of COVID-19 caused by B.1.351 than were unvaccinated people, and had near-total protection from severe disease caused by that strain.

The findings — published on 5 May in *The New England Journal of Medicine*<sup>1</sup> — suggest that current RNA vaccines are a potent weapon against the most worrisome immune-evasive variants. Pfizer, based in New York City, and BioNTech, in Mainz, Germany, are developing an updated RNA vaccine targeting B.1.351, as is Moderna, based in Cambridge, Massachusetts. Early results from Moderna’s efforts suggest that a booster shot of the updated vaccine triggers a strong response against B.1.351.

“I think this variant is probably the worst of all the variants we know,” says Laith Jamal Abu-Raddad, an infectious-disease epidemiologist at Weill Cornell Medicine—Qatar in Doha, who led the Qatari study. “We have the tools, despite these variants, to control at least the severe forms of infection — and this should work quite well on transmission.”

## **Weaker protection**

Researchers in South Africa identified B.1.351 in late 2020, and it’s now the predominant strain there. Laboratory studies show that the variant harbours

mutations that blunt the effects of virus-blocking antibodies, and trials suggest that some COVID-19 vaccines are significantly less effective against the strain than against others.

Early lab research suggested that RNA vaccines, including the Pfizer–BioNTech jab, would be weakened by B.1.351, but probably not fully compromised. In April, the companies announced that a small trial in South Africa had found the vaccine to be fully effective against B.1.351, but the study of 800 people recorded a total of just 6 infections caused by B.1.351 in the placebo group, so efficacy might have been much lower.



### First evidence that COVID vaccines protect people against new variants

Abu-Raddad's team analysed tens of thousands of COVID-19 cases that occurred between the start of Qatar's vaccination campaign in late December and the end of March. Genome sequencing showed that B.1.1.7 and B.1.351 were the predominant coronavirus lineages during this period and, from mid-February, each accounted for about half of the country's cases.

The researchers compared SARS-CoV-2 infection rates in vaccinated people with those in unvaccinated controls. People who received two vaccine doses were about 90% less likely to develop an infection caused by B.1.1.7, echoing findings from Israel, the United Kingdom and elsewhere. The

researchers identified around 1,500 ‘breakthrough’ infections caused by the B.1.351 variant in vaccinated individuals, but only 179 of these occurred more than 2 weeks after the second dose. There were hardly any severe cases of COVID-19 caused by either B.1.1.7 or B.1.351 among fully vaccinated individuals.

“Even though there were breakthrough infections, they didn’t lead to hospitalization and death, except very, very rarely,” says Abu-Raddad. Two people died of COVID-19 caused by B.1.351 after receiving their second vaccine dose, but it is very likely that they were infected before the protective effects of the booster shot began. “If, a year ago, I told somebody we would have 75% effectiveness against the worst variants we had, they would consider this extremely good news,” Abu-Raddad adds.

## Promising data

Shabir Madhi, a vaccinologist at the University of the Witwatersrand in Johannesburg, South Africa, says the Qatari results are promising. The comparatively high levels of virus-blocking antibodies triggered by two doses of an RNA vaccine probably explain why it confers better protection against B.1.351 than do other vaccines, such as the one developed by the University of Oxford, UK, and pharmaceutical company AstraZeneca in Cambridge, UK.



## Massive Google-funded COVID database will track variants and immunity.

But Madhi expects that other vaccines will also prevent severe disease caused by that variant. In another 5 May *New England Journal of Medicine* study<sup>2</sup>, his team reported that the jab produced by biotechnology company Novavax in Gaithersburg, Maryland, lowered the risk of getting COVID-19 by 60% in participants without HIV in a South African trial involving more than 6,000 people. As-yet unpublished data show that the vaccine was highly effective against severe cases of COVID-19 caused by B.1.351, with no cases in vaccinated individuals and five in the placebo arm.

If vaccine efficacy is lower against B.1.351, even highly successful immunization programmes in countries affected by the variant might not reduce cases to the same extent as in countries dealing with less troublesome strains, says Madhi. “Nevertheless, by protecting high-risk individuals, we could still return to a relatively normal lifestyle, even with ongoing circulation.”

Qatar, where more than one-third of the population has received at least one dose of the vaccine, might provide an early glimpse at how the worst coronavirus variants can be controlled. Abu-Raddad says there is evidence that the Pfizer–BioNTech vaccine might also be highly effective at blocking transmission of B.1.351. And after cases of the variant peaked in mid-April, he says, “things have been going extremely well, the numbers are going down very, very rapidly”.

Nature **593**, 325-326 (2021)

doi: <https://doi.org/10.1038/d41586-021-01222-5>

## References

1. 1.

Abu-Raddad, L. J. & Butt, A. A. *N. Engl. J. Med.*  
<https://doi.org/10.1056/NEJMc2104974> (2021).

2. 2.

Shinde, V. *et al.* *N. Engl. J. Med.*  
<https://doi.org/10.1056/NEJMoa2103055> (2021).

## **Jobs from Nature Careers**

- - - [All jobs](#)
    - - **Post-Doc - Numerical and microfabrication development of 4D metamaterials for mechanical, acoustic and photonics applications**

[Université Bourgogne Franche-Comté \(UBFC\)](#)

[BESANCON, France](#)

[JOB POST](#)

  - **Postdoctoral Position - Ecological Modeler**

[The University of British Columbia \(UBC\)](#)

[Vancouver, Canada](#)

[JOB POST](#)

  - **Postdoctoral Fellow | Zandstra Stem Cell Bioengineering Lab**

[The University of British Columbia \(UBC\)](#)

[Vancouver, Canada](#)

[JOB POST](#)

▪ **Personalreferent\*in (d/m/w), Gruppe Personaladministration**

Helmholtz Centre for Heavy Ion Research GmbH (GSI)

Darmstadt, Germany

JOB POST

[Download PDF](#)

[Download PDF](#)

## Related Articles



- [Who received the first billion COVID vaccinations?](#)



- [First evidence that COVID vaccines protect people against new variants](#)



- [Massive Google-funded COVID database will track variants and immunity](#)



- [What Pfizer's landmark COVID vaccine results mean for the pandemic](#)



- **This COVID-vaccine designer is tackling vaccine hesitancy — in churches and on Twitter**
- 

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-021-01222-5>

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

NEWS

06 May 2021

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

Even after mass vaccinations, some hospitalizations and deaths from the coronavirus are inevitable — but opinions differ on how many is too many for a return to relative normality.

- [Smriti Mallapaty](#)
  1. Smriti Mallapaty  
[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)



Londoners enjoy drinks in outside areas of pubs after restrictions are relaxed on 12 April. Credit: Tolga Akmen/AFP/Getty

On 24 April, Perth in Western Australia entered a snap three-day lockdown when two people tested positive for SARS-CoV-2 — the first community infections recorded outside hotel quarantine in the state in more than a year. Pubs, gyms and playgrounds shut, remembrance-day services were cancelled and people were confined to their homes.

Australia is part of a group of countries — including Bhutan, China and New Zealand — that has applied a zero-tolerance approach. When outbreaks are detected, the response is swift and severe: mass testing, sudden lockdowns and closed borders.

But this cannot be sustained indefinitely. “We have to accept that people will get infected, will go to hospital and will die from COVID-19 in the future,” says James McCaw, an infectious-diseases epidemiologist at the University of Melbourne, who advises the Australian government.

As more people get vaccinated, scientists and health officials are pondering how societies can live with the virus, and what level of risk they are willing to absorb. In some countries, such as Australia, the threshold is low. But in some nations worn down by a year of restrictions, such as India and the United States, communities remain open even in the face of high transmission.

## Different nations, different answers

Researchers say there is no universally agreed number of hospitalizations and deaths that societies will find acceptable. But there are some situations that most nations will do their best to avoid, such as excess deaths caused by hospitals being overrun.



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

Annual deaths from diseases such as influenza — which before the pandemic killed between one-quarter of a million and half a million people each year globally — offer a barometer. And in Israel, where vaccination rates are high and life is returning to normal, people seem to have settled on a few deaths a day as the acceptable number, says Eran Segal, a computational biologist at the Weizmann Institute of Science in Rehovot, Israel, who has modelled the pandemic and advises the government.

Scientists and public-health officials are beginning discussions about the acceptable level of risk, but the decisions involve cultural, ethical and political factors, and differ widely between regions. “Each country will set its own threshold,” says Sylvie Briand, head of infectious-hazards management at the World Health Organization in Geneva, Switzerland.

“Every country is almost another experiment in public policy,” says Michael Osterholm, an epidemiologist at the University of Minnesota in Minneapolis.

## Acceptable levels of peril

Researchers are still unclear how annual deaths from COVID-19 will eventually compare with deaths that societies have been willing to endure from influenza and other endemic diseases. In the United Kingdom alone, seasonal flu epidemics kill thousands of people each winter. “This seems to be an acceptable risk to society, and repeated vaccination and some continued distancing measures might keep COVID-19 deaths to this level,” says David Spiegelhalter, a statistician at the University of Cambridge, UK.

But the pandemic has made some societies more risk averse concerning deaths from respiratory infections. In New Zealand, for example, lockdowns almost eliminated the incidence of influenza and respiratory syncytial virus, a common cause of colds. Researchers there are now discussing whether returning to previous numbers of deaths from those pathogens is acceptable, or if more effort should be made to control them.

COVID-19 also presents unique risks that make comparisons with flu difficult. It is a new virus that is deadlier than influenza, which means that its uncontrolled spread could quickly escalate into a devastating situation at hospitals, says Segal.



A mass cremation in India, where COVID-19 cases are surging. Caption: Manish Rajput/SOPA Images/LightRocket/Getty

Societies will also have to consider the impact of '[long COVID](#)', the ongoing symptoms that affect between 10% and 20% of those infected, says Alex James, a mathematical modeller at the University of Canterbury in Christchurch, New Zealand. Even if deaths are low and hospitals can cope, if many people end up with long-term disabilities because of COVID-19, that would be a signal that the rates of infection are too high, she says.

## Stopping hospitals getting overrun

One factor governing what nations will consider an acceptable level of hospitalization and death will be health-system capacity, including the capabilities of intensive-care units (ICUs). "If we have to postpone elective surgeries because our ICUs are full of COVID-19 patients, that's a very poor position," adds James.

Segal estimates that Israel hits that point when 500 ICU beds are filled nationwide. Beyond that precipice, the quality of health care rapidly declines and death rates increase. It's wise to implement a lockdown before that point, he says.

The United Kingdom has followed this general rule throughout the pandemic. There have been three nationwide lockdowns, and each was begun when "it was clear that the number of cases was growing to the point that hospitals were unable to cope", says Graham Medley, an infectious-diseases modeller at the London School of Hygiene & Tropical Medicine, who advises the government.

Israel has among the highest vaccination rates in the world, with 60% of its population having received at least one dose, and it offers a glimpse of what baseline levels of severe disease and deaths might be in a post-vaccination world. The country began opening its economy in February, when about one-third of its population had been fully vaccinated, and numbers of hospitalizations and deaths have continued to decline. In the past few weeks, fewer than 100 new cases have been detected daily, despite high testing rates; a handful of people have been admitted to ICUs each day; and 2–5 people have died.

If those rates are sustained, COVID-19's death toll in Israel could stabilize at 1,000–2,000 a year, says Segal. "Even if those numbers were to rise, nobody would close down the economy now. They would only consider closing if we saw, again, the danger of losing control."

## A changing risk equation

Since the pandemic began, the risk equation — the way people balance the risks from infection against problems caused by harsh restrictions — has changed for many.

Early on, many countries compared the outbreak with the flu pandemic that began in 1918, which killed at least 50 million people, says Briand — "The fear was really intense." But since then, perceptions have changed, as people have balanced the risks from COVID-19 with considerations such as

increased unemployment, she says. Around 3.2 million people are currently known to have died of COVID-19, and the final figure is not expected to be as high as in 1918–20.



### [India's massive COVID surge puzzles scientists](#)

A gruelling year has made some people less willing to adhere to restrictions, and politicians reluctant to impose them. For example, when the B.1.1.7 variant surged in parts of the US Midwest earlier this year, transmission reached levels that had triggered restrictions early in the pandemic — but the response this time was much more muted. “People’s tolerance today is very different,” says Osterholm.

A year ago, the fear of an unknown virus galvanized governments into implementing harsh social restrictions. The risks have since been articulated more clearly and people have begun to factor them into their daily lives, says Medley. This means that people might be more willing than before to put up with the risk of a certain level of hospitalization and death, and that the bar for imposing social restrictions is now higher. But the exact position of that bar is still unknown — and it could come down again if new variants threaten gains from vaccination campaigns, he says.

doi: <https://doi.org/10.1038/d41586-021-01220-7>

## Jobs from Nature Careers

- - - [All jobs](#)
    - 
    - **Post-Doc - Numerical and microfabrication development of 4D metamaterials for mechanical, acoustic and photonics applications**  
[Université Bourgogne Franche-Comté \(UBFC\)](#)  
[BESANCON, France](#)  
[JOB POST](#)
    - **Postdoctoral Position - Ecological Modeler**  
[The University of British Columbia \(UBC\)](#)  
[Vancouver, Canada](#)  
[JOB POST](#)
    - **Postdoctoral Fellow | Zandstra Stem Cell Bioengineering Lab**  
[The University of British Columbia \(UBC\)](#)  
[Vancouver, Canada](#)  
[JOB POST](#)
    -

## Personalreferent\*in (d/m/w), Gruppe Personaladministration

Helmholtz Centre for Heavy Ion Research GmbH (GSI)

Darmstadt, Germany

### JOB POST

Download PDF

Download PDF

## Related Articles



- India's massive COVID surge puzzles scientists



- Has COVID peaked? Maybe, but it's too soon to be sure



- Three million COVID deaths is a grim milestone

---

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-021-01220-7>

## NEWS FEATURE

19 May 2021

# The mysterious microbes that gave rise to complex life

As scientists learn more about enigmatic archaea, they're finding clues about the evolution of the complex cells that make up people, plants and more.

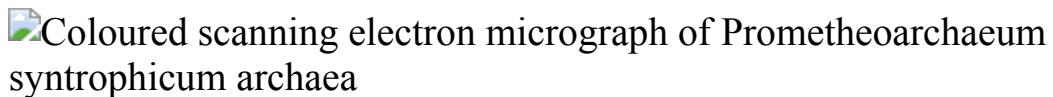
- [Amber Dance](#) <sup>0</sup>

1. Amber Dance

1. Amber Dance is a freelance science journalist in the Los Angeles area of California.

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)



Scientists spent 12 years culturing a slow-growing, tentacled archaeon thought to be similar to the ancestor of complex cells. Credit: Hiroyuki Imachi, Masaru K. Nobu and JAMSTEC

Evolutionary biologist David Baum was thrilled to flick through a preprint in August 2019 and come face-to-face — well, face-to-cell — with a distant cousin. Baum, who works at the University of Wisconsin–Madison, was looking at an archaeon: a type of microorganism best known for living in extreme environments, such as deep-ocean vents and acid lakes. Archaea can look similar to bacteria, but have about as much in common with them as they do with a banana. The one in the bioRxiv preprint had tentacle-like

projections, making the cells look like meatballs with some strands of spaghetti attached.

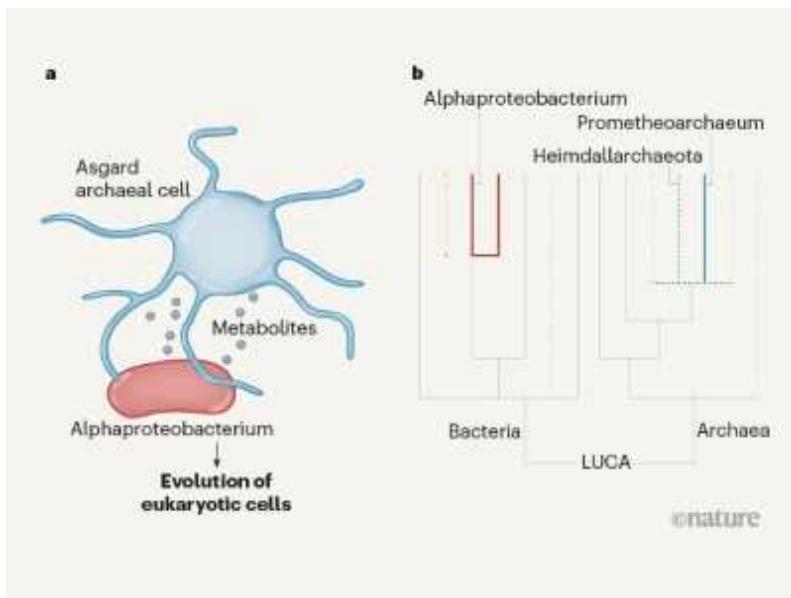
Baum had spent a lot of time imagining what humans' far-flung ancestors might look like, and this microbe was a perfect doppelgänger.

Archaea are more than just oddball lifeforms that thrive in unusual places — they turn out to be quite widespread. Moreover, they might hold the key to understanding how complex life evolved on Earth. Many scientists suspect that an ancient archaeon gave rise to the group of organisms known as eukaryotes, which include amoebae, mushrooms, plants and people — although it's also possible that both eukaryotes and archaea arose from some more distant common ancestor.

Eukaryotic cells are palatial structures with complex internal features, including a nucleus to house genetic material and separate compartments to generate energy and build proteins. A popular theory about their evolution suggests that they descended from an archaeon that, somewhere along the way, merged with another microbe.

But researchers have had trouble exploring this idea, in part because archaea can be hard to grow and study in the laboratory. The microbes have received so little attention that even the basics of their lifestyle — how they develop and divide, for example — remain largely mysterious.

Now, researchers could be closer than ever before to plausible evolutionary answers. Thanks to a surge in interest in these oft-overlooked microbes, and the ongoing invention of [methods for tending to archaea in the lab](#), cell biologists are seeing them in more detail than was previously possible. Publications on this enigmatic group of microbes have nearly doubled over the past decade, and the nascent study of their biology is immensely exciting, says molecular microbiologist Iain Duggin at the University of Technology Sydney in Australia. "We can do some interesting fundamental experiments, and make some major first-step discoveries," he says. "We may be able to get a much clearer view of how the earliest eukaryotes evolved."



### Meet the relatives of our cellular ancestor

The [images that wowed Baum](#), later published in *Nature*, offered such a view<sup>1</sup>. They were the result of 12 years' painstaking culture of an archaeon thought to be closely related to the one that spawned the eukaryotes. Microbiologists worldwide were thrilled by the portraits, but for Baum, they were a pet theory brought to life.

Five years earlier, he and his cousin, cell biologist Buzz Baum at the Medical Research Council (MRC) Laboratory of Molecular Biology (LMB) in Cambridge, UK, had published a hypothesis about the origin of eukaryotes<sup>2</sup>. They predicted that the grandmother of them all might have sprouted protrusions, much like those on the archaeon in the paper. They reasoned that these protrusions came to surround nearby bacteria, which then transformed into a defining feature of eukaryotic cells: the lozenge-shaped energy-makers known as mitochondria.

As David Baum stared at the spaghetti-like strands, he recalls thinking, “Oh my goodness, we were right.”

## Fundamental mysteries

If a eukaryote is really a souped-up archaeon, then scientists must understand archaea to work out how the more-complex cells came to be. Whereas scientists studying eukaryotes and bacteria have been drilling down into processes such as cell division and growth for decades, the inner workings of archaea are still largely obscure. “Archaea, every time, do things differently,” says Sonja Albers, a molecular microbiologist at the University of Freiburg in Germany. For example, related proteins might take on different jobs in different organisms. That makes archaea fascinating to study, says Duggin, but it’s also important, because researchers can then compare across groups, looking for clues to the origin of the nucleus and other major innovations.

From the soils to the seas, one thing all cells have in common is that they split to make more of themselves. It happened in the common ancestor of all cell-based life on Earth, but the process started to look different as organisms adapted to their niches.

Researchers can explore evolution by looking at this divergence. Any mechanisms that all cellular life forms have in common point to biology inherited from the very earliest cells. By contrast, systems shared between only archaea and eukaryotes, or only bacteria and eukaryotes, hint at which parent provided the various ingredients of eukaryote biology. For example, the flexible membrane that separates eukaryotic cells from the outside environment resembles that in bacteria.



Some archaeal species thrive in the scalding water of the Grand Prismatic Spring in Yellowstone National Park in Wyoming. Credit: Getty

Duggin studies cell division in the archaeon *Haloferax volcanii*. It's a lover of salty conditions, such as those in the Dead Sea, and not of volcanoes, as the species moniker suggests. (It was named after microbiologist Benjamin Elazari Volcani.) For an extremophile, *H. volcanii* is pretty simple to grow in a salty broth, and its large, flat cells are easy to see dividing under the microscope.

Despite the enormous differences between bacteria, eukaryotes and archaea, the groups do share a couple of cell-division systems. In bacteria, a protein called FtsZ forms a ring at the future site of cell division. Duggin and his collaborators have observed the same in *H. volcanii*<sup>3</sup>. FtsZ, then, seems to have roots at the very base of the evolutionary tree.

Archaea have helped to surface other ancient proteins, too. One is SepF, a protein that Albers's group has found is essential to *H. volcanii* division<sup>4</sup>. Together with FtsZ, it could be part of a primordial "minimal system" for cell division, according to Nika Pende, an evolutionary biologist at the

Pasteur Institute in Paris. Pende has analysed the distribution of the genes encoding FtsZ and SepF across a variety of microbes and traced them all the way back to the last universal common ancestor of all living cells<sup>5</sup>.

Yet, at some point in evolution, some archaea assigned the cell-division job to a different set of proteins. This is where Buzz Baum's latest work comes in. His group has been studying the archaeon *Sulfolobus acidocaldarius*. In this case, the name fits: it loves acid and heat. Lab members wear gardening gloves to protect themselves from the acidic liquid it lives in, and built a special chamber so they could watch it divide under the microscope without cool spots or evaporation.



Scientists are studying how archaea such as *Sulfolobus* (left), *Halobacterium* (middle) and *Methanosaarcina* (right) grow and divide to shed light on the evolution of complex cells. Credit: (Left) Eye of Science/SPL; (Centre and Right) Denis Kunkel Microscopy/SPL

Baum's team saw a completely different group of proteins managing the division ring. In eukaryotes, where they were first discovered, these proteins aren't just involved with division. They have a much broader role, pinching membranes apart all over the cell to create membrane-wrapped packages called vesicles, and other small containers. The proteins are known as ESCRTs (endosomal sorting complexes required for transport). In *S. acidocaldarius*, the team saw archaeal proteins related to these all-purpose

pinchers managing the division ring<sup>6</sup>, suggesting that early versions of ESCRTs evolved in the archaeal ancestor of eukaryotes.

FtsZ, meanwhile, evolved into eukaryotic tubulin, which gives structure to our cells. These discoveries suggest that the archaeal ancestor of eukaryotes probably had a kit for shaping and dividing cells that natural selection then adapted to the needs of the more complex descendant cells.

## Glimpsing grandmother

But what kind of cell was that ancestor archaeon? And how did it meet, and merge with, its bacterial partners?

Biologist Lynn Margulis was the first to propose, in 1967, that eukaryotes arose when one cell swallowed others<sup>7</sup>. Most researchers agree that some engulfment went on, but they have different ideas about when that happened, and how the internal compartments in eukaryotes came about. “Several dozen models that were tested have died along the way because they’re no longer plausible,” says Sven Gould, an evolutionary cell biologist at Heinrich Heine University in Düsseldorf, Germany. Other theories might rise or fall as cell biologists add to their understanding of archaea.

Many models assume that the cells that eventually became eukaryotic were already quite complex, with flexible membranes and internal compartments, before they ever met the bacterium that was to become the mitochondrion. These theories require cells to have developed a way of gobbling up external material, known as phagocytosis, so they could snap up the passing bacterium in a fateful bite (see ‘Two ways to make complex cells’). By contrast, Gould and others think that mitochondria were acquired early on, and that they then helped to fuel a larger, more complex cell.

The Baums’ model is one of few to explain how mitochondria could arise without phagocytosis. David Baum first came up with the idea as an undergraduate at the University of Oxford, UK, in 1984. His process starts with archaea and bacteria hanging out, sharing resources. The archaeon might start to stretch and bulge its exterior membranes to boost the surface area for nutrient exchange. With time, those bulges might spread and grow

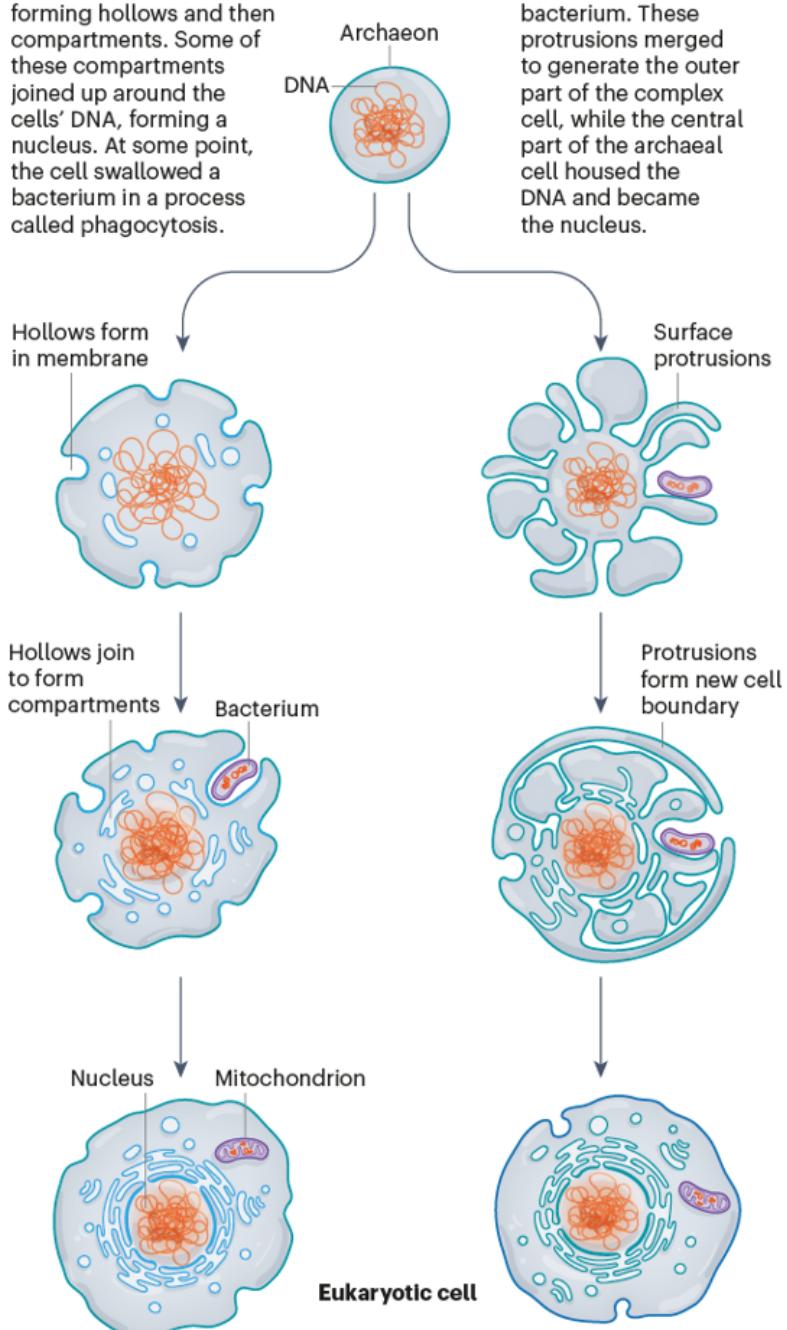
around the bacteria until the bacteria were, more or less, inside the archaeon. At the same time, the archaeon's original exterior membrane, now dwarfed by the long tentacles surrounding it, would evolve into the boundary of the new nucleus, while the cell's new exterior membrane would form when some particularly long tentacles grew right around the edge, greatly enlarging the cell compared to its archaeal precursor. This process differs from phagocytosis, in that it starts with a community of organisms and takes place over long timescales, rather than in a single bite.

## TWO WAYS TO MAKE COMPLEX CELLS

Many researchers think that the cells of eukaryotes — organisms whose cells have complex internal structures — evolved when a bacterium merged with a type of microbe known as an archaeon. Over time, the bacteria became mitochondria, the energy-producing modules inside eukaryotic cells. But how did the union take place?

One leading theory suggests that the membranes of ancient archaea folded inwards, forming hollows and then compartments. Some of these compartments joined up around the cells' DNA, forming a nucleus. At some point, the cell swallowed a bacterium in a process called phagocytosis.

A newer idea proposes that an archaeal cell grew protrusions that surrounded a bacterium. These protrusions merged to generate the outer part of the complex cell, while the central part of the archaeal cell housed the DNA and became the nucleus.



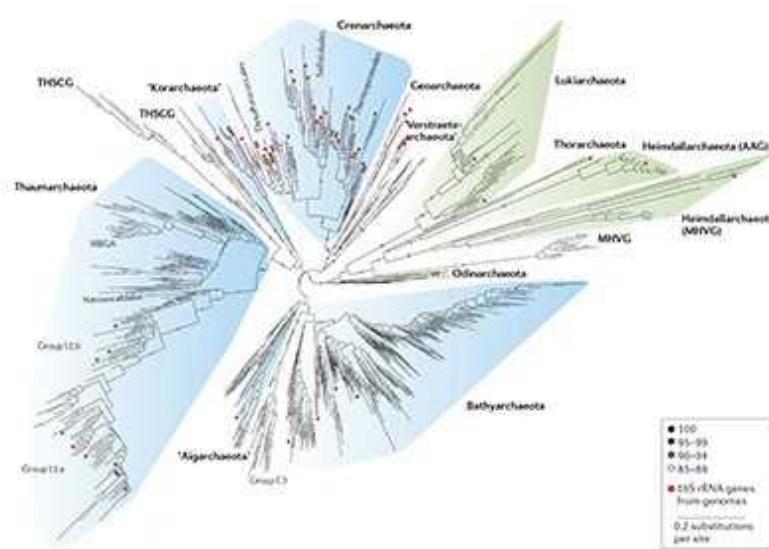
Nik Spencer/*Nature*; Source: B. Baum & D. A. Baum *BMC Biol.* **18**, 72 (2020).

David Baum's tutor told him the idea was creative, but lacking in evidence. He set it aside. But he'd already shared his enthusiasm for life science with his cousin Buzz, a child then, at regular family dinners in Oxford. "That's partly why I went into biology," recalls Buzz.

In 2013, David decided to write up his theory. He sent a note to Buzz, by now running his own lab, who helped develop the theory further. The duo defined several aspects of biology that support their idea, such as the fact that archaea and bacteria have been found living side by side and trading nutrients. The Baums struggled to publish their proposal, but it finally found a home at *BMC Biology*<sup>2</sup> in 2014.

The idea received an enthusiastic response, Buzz recalls, especially from cell biologists. But back in 2014, David still thought they had just a 50–50 chance of being right.

And then, five years later, the spaghetti-and-meatball images appeared. Both Baums were thrilled.



## Archaea and the origins of eukaryotes

The species was the first to be cultured from a group called the Asgard archaea. These organisms, described in 2015, have genes encoding proteins that many scientists consider remarkably similar to those of eukaryotes<sup>8</sup>. Researchers quickly came to suspect that the archaeal ancestor of eukaryotes was something akin to an Asgard archaeon. By pointing to a potential grandmother, the discovery supported the Baums' hypothesis.

The Asgard representative — which doesn't yet have a finalized name, and is currently known as *Candidatus 'Prometheoarchaeum syntrophicum'* — grew in a bioreactor alongside either of a pair of microbial hangers-on with which it shared nutrients. Notably, it lacked any complex internal membranes or signs that it could ever hope to phagocytose those associates. It had three systems that could be associated with cell division: proteins that are equivalent to FtsZ; ESCRTs; and the muscle-contraction protein actin, which also contributes to division in eukaryotes. The culturers haven't yet worked out which it uses to split itself, says team member Masaru Nobu, a microbiologist at the National Institute of Advanced Industrial Science and Technology in Tokyo.

The big surprise came when the cells stopped dividing and sprouted tentacles. It's possible, the Baums suggest, that these might amplify nutrient exchange with the microbes that the archaeon was co-cultured with, as their model predicted for the grandmother cell.

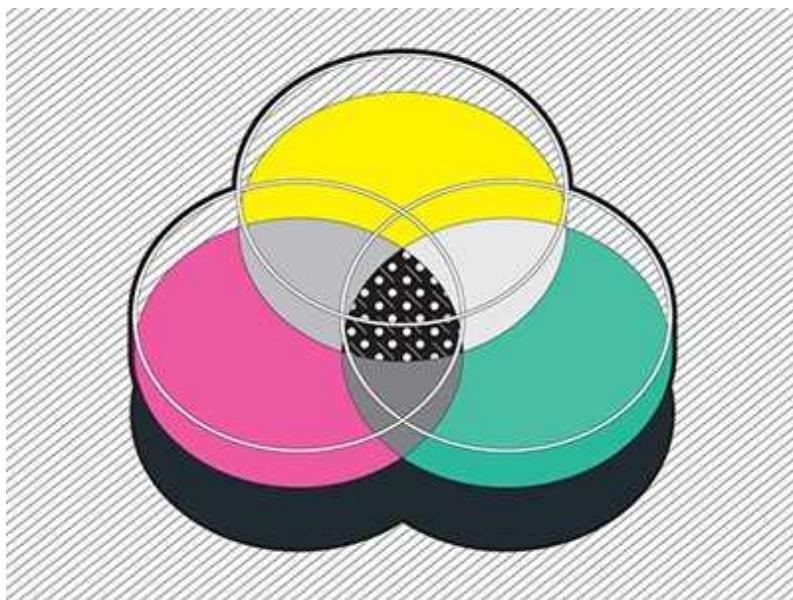
On the basis of their observations, Nobu and his colleagues developed a theory about how eukaryotes evolved that shares much with the Baums' idea. It involves one microbe extending filaments that eventually engulf its partner<sup>1</sup>. "I like our hypothesis because it allows for these complexities that are unique to eukaryotes" — nuclei and mitochondria — "happening at the same time", says Nobu.

## Culturing confidence

The pictures of the Asgard archaeon really helped to shore up the Baums' theory. "It's very exciting that they form these protrusions," says evolutionary microbiologist and Asgard co-discoverer Anja Spang at the NIOZ Royal Netherlands Institute for Sea Research on the island of Texel.

“It all ties together, because if an ancestor could form such protrusions, it could make a consortium of archaea and bacteria a lot more tight.”

The Baums now estimate there’s an 80% chance they’re on the right track, and they’re not the only ones gaining confidence. Ramanujan Hegde, a biochemist at the LMB who studies membrane proteins, is contributing to the upcoming seventh edition of the textbook *Molecular Biology of the Cell*. He and his colleagues decided that the Baum hypothesis will replace the phagocytosis-based model in the current edition. But there’s still no proof, of course: Hegde is careful to use uncertain terms such as “could have”.



### Mining the microbial dark matter

Indeed, some others, including Gould, say the Baums’ model doesn’t fully explain how those membrane protrusions could have evolved into sheets, closed around the cell to create a complete outer boundary or acquired the characteristics of bacterial membranes. To explain the bacteria-like membranes, Gould and his colleagues have developed a model based on the fact that both free-living bacteria and mitochondria regularly release vesicles. They proposed in 2016 that the proto-eukaryote first acquired mitochondria — their theory doesn’t specify how — which oozed vesicles into the cell. These vesicles provided the membrane materials that the evolving eukaryotic cell used to build its inner structure and external

border<sup>9</sup>. This would explain why eukaryotes' membranes look like bacteria's, says Gould.

These and other competing models could be either supported or refuted as researchers continue to culture and study archaea; dozens of the microbes have now been grown successfully in the lab. Buzz Baum and his collaborators are investigating symbiosis in archaea and analysing microbial family trees to test their idea further. Nobu and his colleagues are investigating the protrusions in more detail and working on other Asgard archaea.

There might be more evidence waiting to be found. For example, the Baums predict that it might be possible to discover eukaryotes in which the tentacle membranes haven't quite disconnected from the exterior cell membrane, corresponding to an intermediate in their theory. What's looking more and more likely, at least, is that we owe our existence to an ancient love story of sorts between an archaeon and a bacterium. "We are part bacteria, part archaea, part new inventions," says Buzz Baum. "It's better together."

Nature **593**, 328-330 (2021)

doi: <https://doi.org/10.1038/d41586-021-01316-0>

## References

1. 1.  
Imachi, H. *et al. Nature* **577**, 519–525 (2020).
2. 2.  
Baum, D. A. & Baum, B. *BMC Biol.* **12**, 76 (2014).
3. 3.  
Liao, Y. *et al. Nature Microbiol.* **6**, 594–605 (2021).
4. 4.

Nußbaum, P., Gerstner, M., Dingethal, M., Erb, C. & Albers, S. V.  
Preprint at bioRxiv <https://doi.org/10.1101/2020.10.06.327809> (2020).

5. 5.

Pende, N. *et al.* Preprint at bioRxiv  
<https://doi.org/10.1101/2020.10.06.328377> (2020).

6. 6.

Pulschen, A. A. *et al.* *Curr. Biol.* **30**, 2852–2859 (2020).

7. 7.

Sagan, L. *J. Theor. Biol.* **14**, 255–274 (1967).

8. 8.

Spang, A. *et al.* *Nature* **521**, 173–179 (2015).

9. 9.

Gould, S. B., Garg, S. G. & Martin, W. F. *Trends Microbiol.* **24**, 525–534 (2016).

## Jobs from Nature Careers

- - - [All jobs](#)
    - 
    - [\*\*Post-Doc - Numerical and microfabrication development of 4D metamaterials for mechanical, acoustic and photonics applications\*\*](#)

[Université Bourgogne Franche-Comté \(UBFC\)](#)

[BESANCON, France](#)

[JOB POST](#)

- [\*\*Postdoctoral Position - Ecological Modeler\*\*](#)

[The University of British Columbia \(UBC\)](#)

[Vancouver, Canada](#)

[JOB POST](#)

- [\*\*Postdoctoral Fellow | Zandstra Stem Cell Bioengineering Lab\*\*](#)

[The University of British Columbia \(UBC\)](#)

[Vancouver, Canada](#)

[JOB POST](#)

- [\*\*Personalreferent\\*in \(d/m/w\), Gruppe Personaladministration\*\*](#)

[Helmholtz Centre for Heavy Ion Research GmbH \(GSI\)](#)

[Darmstadt, Germany](#)

[JOB POST](#)

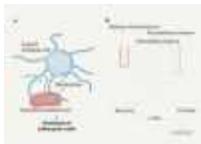
[Download PDF](#)

[Download PDF](#)

## **Related Articles**



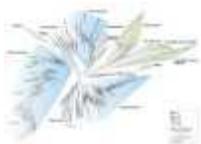
- [The trickster microbes that are shaking up the tree of life](#)



- [Meet the relatives of our cellular ancestor](#)



- [Mining the microbial dark matter](#)



- [Archaea and the origins of eukaryotes](#)

---

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-021-01316-0>

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

## NEWS FEATURE

18 May 2021

# How COVID is changing the study of human behaviour

The pandemic is teaching us key lessons about crisis, communication and misinformation, and is spurring changes in the way scientists study public-health questions.

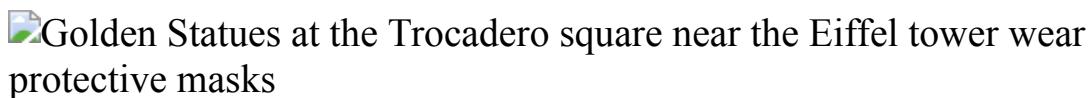
- [Christie Aschwanden](#) <sup>0</sup>

1. Christie Aschwanden

1. Christie Aschwanden is a science journalist in Cedaredge, Colorado.

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)



National identity plays a part in how likely people are to support public-health policies such as mask wearing. Credit: Benoit Tessier/Reuters/Alamy

During the early months of the COVID-19 pandemic, Jay Van Bavel, a psychologist at New York University, wanted to identify the social factors that best predict a person's support for public-health measures, such as physical distancing or closing restaurants. He had a handful of collaborators ready to collect survey data. But because the pandemic was going on everywhere, he wondered whether he could scale up the project. So he tried something he'd never done before.

He posted a [description of the study on Twitter](#) in April, with an invitation for other researchers to join. “Maybe I’ll get ten more people and some more data points,” he recalls thinking at the time. Instead, the response floored him. More than 200 scientists from 67 countries joined the effort. In the end, the researchers were able to collect data on more than 46,000 people. “It was a massive collaboration,” he says. The team showed how, on the whole, people who reported that national identity was important to them were more likely to support public-health policies<sup>1</sup>. The work is currently being peer reviewed.

For social scientists, the COVID-19 pandemic has presented a unique opportunity — a natural experiment that “cuts across all cultures and socio-economic groups”, says Andreas Olsson, a psychologist at the Karolinska Institute in Stockholm. Everyone is facing similar threats to their health and livelihoods, “so we can see how people respond differently to this depending on culture, social groups and individual differences”, he says. Researchers have been able to compare people’s behaviours before and after large policy changes, for example, or to study the flow of information and misinformation more easily.

The pandemic’s global scope has brought groups together from around the world as never before. And with so much simultaneous interest, researchers can test ideas and interventions more rapidly than before. It has also forced many social scientists to adapt their methods during a time when in-person interviews and experiments have been next to impossible. Some expect that innovations spurred by the pandemic could outlive the current crisis and might even permanently change the field.

For example, with the technology that’s now tried and tested, Van Bavel says, it’s much easier to build an international team. “Now that we’ve got the infrastructure and experience, we’ll be able to do this for all kinds of things,” he says.

## Social vaccine boosters

Before Van Bavel’s massive collaboration, he and a group of more than 40 researchers got together to outline the ways in which behavioural research

might inform and improve the response to the SARS-CoV-2 coronavirus at a time when people are scared, sceptical and inundated by information. They outlined previous research in the field that might influence policy, and identified potential projects on threat perception, decision-making and science communication, among other things<sup>2</sup>.



### The race to curb the spread of COVID vaccine disinformation

Many were eager to apply their work towards understanding the public response to practices such as lockdowns and mask mandates. In the survey of more than 46,000 people, Van Bavel and his colleagues showed that countries in which people were most in favour of precautionary measures tended to be those that fostered a sense of public unity and cohesion. A sense, he says, that “we’re all in this together”. That was somewhat counter-intuitive. Right-wing political ideology correlated with resistance to public-health measures among survey participants, but, on the whole, a strong national identity predicted more support for such measures. Van Bavel says this suggests that it might be possible to leverage national identity when promoting public-health policies.

Other work has shown that who delivers the message really matters. A study<sup>3</sup> published in February surveyed more than 12,000 people in 6 countries — Brazil, Italy, South Korea, Spain, Switzerland and the United States — about their willingness to share a message encouraging social

distancing. The message could be endorsed by actor Tom Hanks, celebrity Kim Kardashian, a prominent government official from the survey-taker's country or [Anthony Fauci](#), director of the US National Institute of Allergy and Infectious Diseases in Bethesda, Maryland. Respondents from all countries were most willing to share the message when it came from Fauci (although in the United States, where COVID-19 has been highly politicized, he has become a divisive figure for some). Celebrity endorsements were relatively ineffective by comparison.

Preliminary research suggests that aligning the message with recipients' values or highlighting social approval can also be influential. Michele Gelfand, a psychologist at the University of Maryland in College Park, is part of a team running an 'intervention tournament' to identify ways of promoting mask wearing among conservatives and liberals in the United States.



A London billboard encouraged people to follow lockdown guidance to prevent COVID-19. Credit: May James/SOPA Images/Zuma

The researchers are testing eight interventions, or ‘nudges’, that reflect different moral values and factors specific to COVID-19. The aim is to work out which are most effective at encouraging these political groups to adhere to public-health guidance. One message they are testing emphasizes that mask wearing will ‘help us to reopen our economy more quickly’ — an approach designed to appeal to Republicans, who are [more likely to view the pandemic as an economic crisis than a health one](#). Another intervention highlights harm avoidance — a value that liberal people say is important to them. The message emphasizes that a mask ‘will keep you safe’.

“We’re pitting them against one another to see which nudge works best,” Gelfand says. It’s a study design that can test multiple interventions simultaneously, and could be deployed on a large scale across many geographical regions — a benefit made more urgent by the pandemic. The results have not yet been published.

Others started using a similar approach to encourage vaccination even before a SARS-CoV-2 vaccine was available. The Behavior Change For Good Initiative at the University of Pennsylvania in Philadelphia was testing nudges that encourage people to get the influenza vaccine. Katherine Milkman, a behavioural researcher at the university’s Wharton School, and her colleagues tested around 20 messaging strategies — everything from jokes to direct appeals. “We’re seeing things that work,” Milkman says. They’ve found, for example, that texting people to say a flu shot had been reserved especially for them boosted vaccination rates<sup>4,5</sup>.



## Tracking QAnon: how Trump turned conspiracy-theory research upside down

The findings were almost immediately put to work by researchers seeking to increase COVID-19 vaccination uptake. Researchers at the University of California, Los Angeles (UCLA), tried replicating the strategy among people being treated at the UCLA Health system in February and March<sup>6</sup>, and found that it “proved quite useful for nudging COVID-19 vaccination”, Milkman says.

And, in March, Milkman received an e-mail from Steve Martin, chief executive of the behavioural-science consultancy Influence at Work in Harpenden, UK, telling her that his team had implemented her findings on the island of Jersey in the English Channel. Martin and his colleague Rebecca Sherrington, associate chief nurse for the Government of Jersey, incorporated Milkman’s insight that it was possible to increase the likelihood of someone coming in for a vaccine if they were given “a sense of ownership” — for instance, by telling them that ‘this vaccine has been reserved for you’. “We’ve had a real problem engaging care-home staff — particularly young females, many of whom are sceptical about the vaccine,” Martin says. But using Milkman’s approach, along with other insights (such as the idea that the messenger’s identity also matters), Martin’s programme attained 93% coverage of care-home staff on Jersey, compared with around 80% in other jurisdictions.

## **Depolarization research**

Technologies such as geotracking are helping social scientists to trace the way people really behave, not just how they say they do. The response to the COVID-19 pandemic has shown a dramatic split along political lines in many places, and because so many people own smartphones that include GPS trackers, researchers can quantify how partisanship has translated into behaviour during the pandemic.

Van Bavel and his colleagues used geotracking data from 15 million smartphones per day to look at correlations between US voting patterns and adherence to public-health recommendations<sup>7</sup>. People in counties that voted

for Republican Donald Trump in the 2016 presidential election, for example, practised 14% less physical distancing between March and May 2020 than did people in areas that voted for Democrat Hillary Clinton. The study also identified a correlation between the consumption of conservative news and reduced physical distancing, and found that the partisan differences regarding physical distancing increased over time.



### The epic battle against coronavirus misinformation and conspiracy theories

The research possibilities opened up by geotracking are “beyond my dreams”, says Walter Quattrociocchi, a data scientist at the Ca’Foscari University of Venice, Italy. “We have so much more data to measure social processes now,” he says, and the pandemic has provided a way to put these data to work.

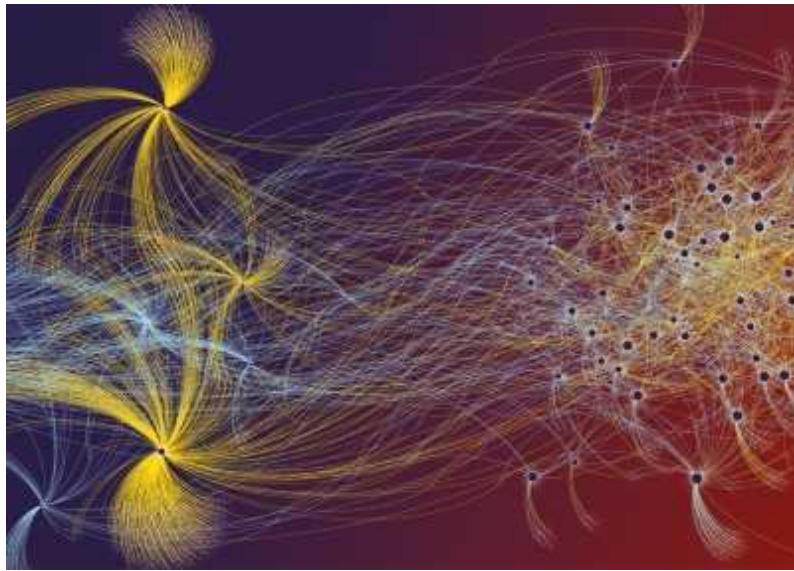
His group used location data from 13 million Facebook users to look at how people moved around France, Italy and the United Kingdom during the early months of the pandemic. The three countries displayed different patterns of mobility that reflect their underlying infrastructure and geography. Movements in the United Kingdom and France were more centralized around London and Paris, respectively, but were more dispersed among Italy’s major population centres<sup>8</sup>. Such results, he says, could help to predict economic resilience in the face of other disasters.

Researchers are also increasingly using Internet-based surveys, a trend accelerated by the pandemic. A US study<sup>9</sup> of people's daily activities during the pandemic — such as going to work, visiting family or dining at restaurants — received more than 6,700 responses per day on average. Results showed that political partisanship had a much greater role than did local COVID-19 rates in influencing safe behaviours. Self-identified Republicans were nearly 28% more likely to be mobile than Democrats were, and this gap widened over the course of the study period from April to September last year.

## Post-lockdown legacy

The pandemic is clearly changing how researchers study behaviour — and in ways that could outlast the lockdowns. "I think people will continue to seek to do bigger studies with more laboratories to produce more robust and widely applicable findings," says Van Bavel. The samples collected through these projects are more diverse than they are for typical approaches, and so the impact from these studies could be much higher, he says.

The COVID-19 crisis has also made researchers much more willing to collaborate and share information, says Milkman. And the pace of publishing and implementing findings has sped up, she says. "I wrote a paper about some of our findings over the Christmas holidays in a week," she says — work that would have normally taken her several months. She expedited the manuscript because she felt the findings were urgent and she wanted to get them into the public domain.



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

The constraints of COVID-19 have nudged social science in a good direction, says Milkman. “We should be doing ‘big science’,” she says, in the way that fields such as physics and astronomy do. Instead of running single, small experiments, researchers can now conduct mega-studies that bring together large groups of researchers to test 20 or even 50 treatment arms at once, she says.

The inability to gather people indoors to conduct research has also forced innovations in how scientists recruit and study participants, says Wändi Bruine de Bruin, a behavioural scientist at the University of Southern California in Los Angeles. She is an investigator on the Understanding America Study, which has been repeatedly surveying about 9,000 nationally representative US households on questions related to the pandemic, such as ‘Do you intend to get vaccinated?’ and ‘How likely do you think it is that you will become infected?’. Being forced to develop procedures to recruit large, nationally representative samples has allowed Bruine de Bruin and her colleagues to recruit more widely. “You don’t have to stay local,” she says, and because participants don’t have to come into the lab, it’s easier to recruit a more diverse sample. “I do think it will push social science forward,” she says.

Technical workarounds spurred by the pandemic could also end up strengthening science. Alexander Holcombe, a psychologist at the University of Sydney, Australia, studies visual perception, which he describes as “a very narrow area of science where people weren’t doing online studies” before the pandemic. Social-distancing practices forced him and his team to learn the computer programming necessary to make their experiments work online. The upshot is that they’re able to get bigger sample sizes, he says — an important improvement on the methodology.

Brian Nosek, executive director at the Center for Open Science, a non-profit organization in Charlottesville, Virginia, sees the pandemic as a chance to rethink some of the fundamentals of how science is done. “It’s given us an occasion to say, ‘Well, how should we be doing this?’” he says, with ‘this’ being everything from teaching and lab work, to study designs and collaboration. The ways that people communicate in the field and engage with collaborators have “fundamentally changed”, he says. “I don’t imagine we’ll go back.”

Nature **593**, 331-333 (2021)

doi: <https://doi.org/10.1038/d41586-021-01317-z>

## References

1. 1.

Van Bavel, J. J. *et al.* Preprint at PsyArXiv  
<https://doi.org/10.31234/osf.io/ydt95> (2020).

2. 2.

Van Bavel, J. J. *et al.* *Nature Hum. Behav.* **4**, 460–471 (2020).

3. 3.

Abu-Akel, A., Spitz, A. & West, R. *PLoS ONE* **16**, e0245100 (2021).

4. 4.

Milkman, K. L. *et al.* Rotman School of Management Working Paper no. 3780267; preprint at SSRN <https://doi.org/10.2139/ssrn.3780267> (2021).

5. 5.

Milkman, K. L. *et al.* Preprint at SSRN  
<https://doi.org/10.2139/ssrn.3780356> (2021).

6. 6.

Dai, H. *et al.* Preprint at SSRN <https://doi.org/10.2139/ssrn.3817832> (2021).

7. 7.

Gollwitzer, A. *et al.* *Nature Hum. Behav.* **4**, 1186–1197 (2020).

8. 8.

Galeazzi, A. *et al.* Preprint at <https://arxiv.org/abs/2005.06341> (2020).

9. 9.

Clinton, J., Cohen, J., Lapinski, J. & Trussler, M. *Sci. Adv.* **7**, eabd7204 (2021).

## Jobs from Nature Careers

- - - [All jobs](#)
    - - [Post-Doc - Numerical and microfabrication development of 4D metamaterials for mechanical, acoustic and photonics applications](#)

Université Bourgogne Franche-Comté (UBFC)

BESANCON, France

JOB POST

▪ **Postdoctoral Position - Ecological Modeler**

The University of British Columbia (UBC)

Vancouver, Canada

JOB POST

▪ **Postdoctoral Fellow | Zandstra Stem Cell Bioengineering Lab**

The University of British Columbia (UBC)

Vancouver, Canada

JOB POST

▪ **Masterthesis / internship (m/f/x)**

Helmholtz Centre for Environmental Research (UFZ)

Leipzig, Germany

JOB POST

Download PDF

Download PDF

## **Related Articles**



- [The race to curb the spread of COVID vaccine disinformation](#)



- [The epic battle against coronavirus misinformation and conspiracy theories](#)



- [Tracking QAnon: how Trump turned conspiracy-theory research upside down](#)



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

---

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-021-01317-z>

## Books & Arts

- **The singing neutrino Nobel laureate who nearly bombed Nevada** [ 17 May 2021]

Book Review • From desert to gold mine — Frederick Reines was a larger-than-life physicist who did larger-than-life experiments.

## BOOK REVIEW

17 May 2021

# The singing neutrino Nobel laureate who nearly bombed Nevada

From desert to gold mine — Frederick Reines was a larger-than-life physicist who did larger-than-life experiments.

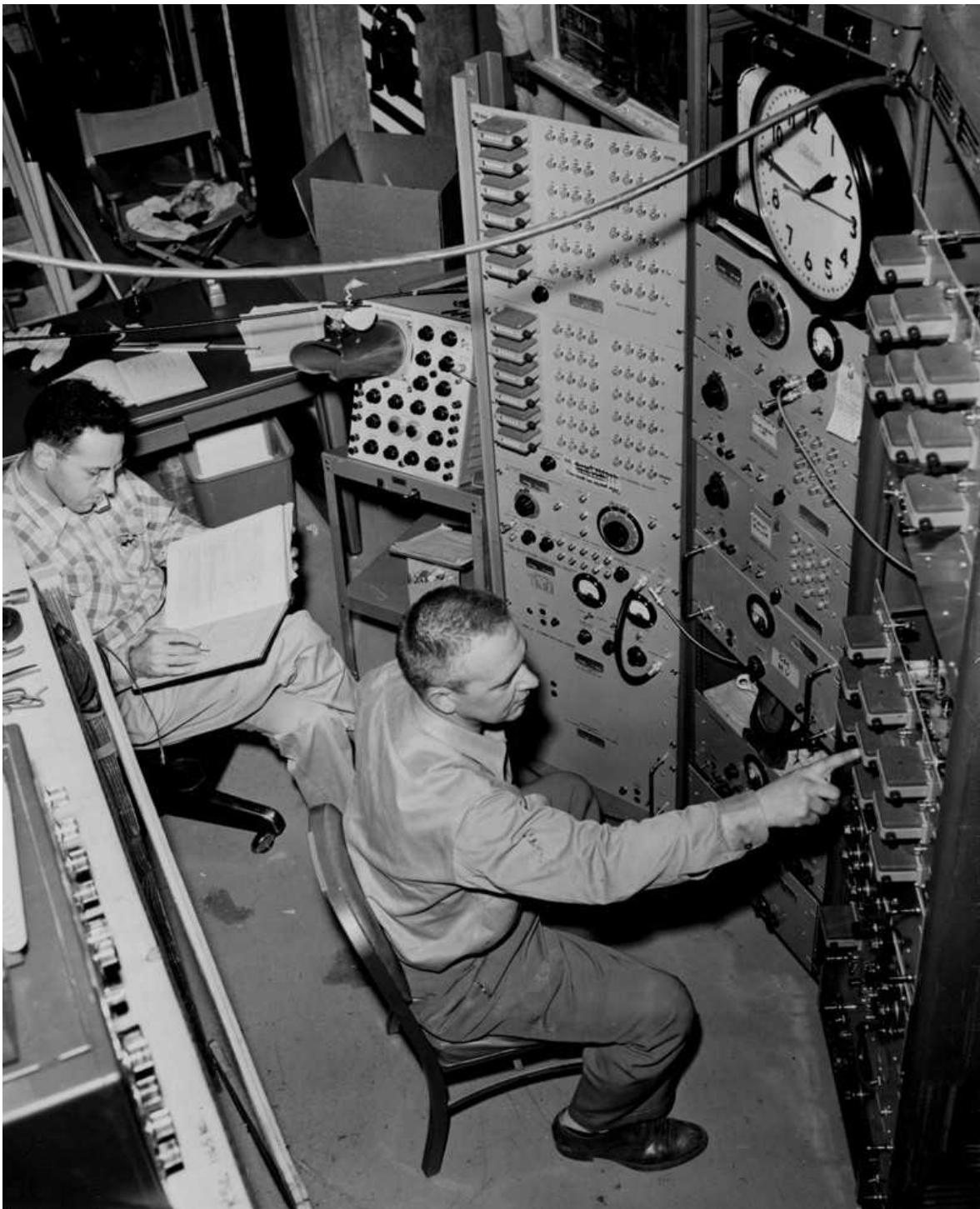
- [Alison Abbott](#) 0

1. [Alison Abbott](#)

1. Alison Abbott is a science writer based in Munich, Germany.

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)



Frederick Reines, left, and Clyde Cowan, at the controls of the Savannah River experiment, where they were the first researchers to detect neutrinos, in 1956. Credit: Los Alamos National Laboratory

**Chasing the Ghost: Nobelist Fred Reines and the Neutrino** *Leonard A. Cole* World Scientific (2021)

In the early 1950s, the physicist Frederick Reines and his colleague Clyde Cowan designed an experiment to detect neutrinos, the tiniest and most elusive of subatomic particles. Theorists were convinced that neutrinos must exist — and that they would be untraceable. And Reines liked nothing better than a challenge.

The experiment was to take place in the Nevada desert. A flux of neutrinos would be created by detonating a 20-kiloton nuclear bomb, comparable to that dropped on Hiroshima, Japan, a few years earlier. A deep hole would be dug 40 metres away from the detonation site, into which a detector would be dropped at the moment of explosion to catch the flux at its maximum.

Eminent physicists enthused about the plan. It was approved by Reines's employer, the government-funded Los Alamos laboratory in New Mexico. Work began on the detector, nicknamed El Monstro, and on the construction of the shaft. At the last minute, Reines and Cowan transferred the experiment to a nuclear reactor, but not because of environmental or safety concerns. They had worked out that although the reactor would deliver a flux of neutrinos three orders of magnitude lower than that from the bomb, it offered a better option for distinguishing signal from noise.

So they did the work instead at the Savannah River nuclear reactor in South Carolina, and Reines and Cowan became the first scientists to detect neutrinos. In 1995, Reines won a share of the Nobel Prize in Physics. (Cowan had died by then.)



### Frederick Reines (1918-98)

The idea of including a nuclear bomb in a basic-research protocol might sound outlandish, but in *Chasing the Ghost*, his biography of Reines, Leonard Cole reminds us that attitudes were different then. He also reminds us of this when describing Reines's work on nuclear bombs during and after the Second World War, and his creation of a neutrino laboratory deep in a gold mine in South Africa in the 1960s, in defiance of academic sanctions against the apartheid state. (However, Cole is hardly an independent voice, being Reines's admiring younger cousin.)

The neutrino-research community has mushroomed over the decades, as it has become clear that these elementary particles are key to understanding the physics of the Universe. Reines was probably its most rambunctious member. Cole relies on many written sources in his reconstruction of Reines's life. He also interviewed scientists, many now in their nineties, who worked with him. He builds a picture of a larger-than-life figure who conducted larger-than-life experiments. The man who rises off the page is an inspiring, supportive colleague and an entertainingly boisterous companion, who whistled and sang his way through life.

## **Los Alamos**

Reines was born in New Jersey in 1918, into a family of Eastern European Jewish immigrants. He was a self-confident high achiever from his youth, declaring in his high-school yearbook his ambition “to be a physicist extraordinary”. He was also a gifted gymnast and musician, who considered a career as an opera singer.

As the Second World War raged, he studied the newly identified phenomenon of nuclear fission, a subject so sensitive that publication of his PhD thesis was delayed until after the war. He joined the Los Alamos laboratory in 1944 to work on the Manhattan Project, the US effort to develop an atomic bomb led by some of the world’s top physicists. Most left Los Alamos after the war; not Reines. He continued to work on the radiation emissions of nuclear bombs — above-ground atomic tests were still taking place, in Nevada and on remote atolls in the Pacific Ocean — and on the development of the even more powerful hydrogen bomb.



Frederick Reines (far right) with Clyde Cowan (far left) and other members of Project Poltergeist, their attempt to detect neutrinos. Credit: LANL/Science Source/Science Photo Library

After the success of the Manhattan Project, the Los Alamos laboratory continued to be well funded, and investigations into fundamental physics were encouraged. Reines's neutrino fixation began when he took a year's sabbatical from his daily responsibilities and found a like-minded colleague in Cowan.

Neutrinos were nicknamed ghost particles because of their uncanny properties. They are the most abundant particles in the Universe — around 100 trillion pass harmlessly through your body every second. They are created in many natural processes, including nuclear reactions in the Sun, explosions of stars and processes that generate radioactivity in Earth's core. They are also made by particle accelerators and nuclear power plants.

But with no charge and a vanishingly small mass, they can be detected only indirectly, when they interact with another particle. Detectors are made from liquids that generate a suitable signal, such as a flash of light, during interactions, and electronics that convert the signal to an electrical pulse. Neutrino interactions are exceedingly rare, however, so the detectors have to be very large, and physicists very patient.

## Project poltergeist

As Cole describes, Reines and Cowan began what they named Project Poltergeist at a plutonium-producing reactor in Hanford, Washington. The results were inconclusive. Undeterred, they transported their bulky yet delicate equipment — including a detector tank containing more than 1 tonne of scintillation fluid — in a crawling convoy of 5 oversized trucks, travelling 2,600 kilometres to the more powerful Savannah River reactor, which they expected to deliver a larger flux of neutrinos. In June 1956, they declared their success.

Reines left Los Alamos in 1959 for the Case Institute in Cleveland, Ohio. Seven years later, he moved to the University of California, Irvine. At Case,

he began a search for natural-source neutrinos, for which he had to build an underground laboratory, sheltered from interfering cosmic radiation. He targeted the East Rand gold mine near Johannesburg, South Africa, which coiled more than 3.5 kilometres below Earth's surface.



### [Neutrinos could shed light on why the Universe has so much more matter than antimatter](#)

Cole describes the extraordinary effort to create a lab there, and the harsh and dangerous conditions for the miners, who — unlike the scientists — wore no protective gear beyond a hard hat. It took six months of blasting to create the required horizontal space. They had to transport newly built equipment from the Case Institute, including 20 tonnes of scintillation fluid in 50 containment tanks, which proved a logistical nightmare. But by 1965, the lab was detecting atmospheric neutrinos.

It made diplomatic waves. The apartheid regime in South Africa was widely distrusted. Some other African countries alleged that it intended to test nuclear weapons underground. As that issue faded, problems arose at home, writes Cole. In 1968, colleagues at Irvine openly challenged academic collaboration in an explicitly racist country. Reines countered that, in his view, science transcended politics. Still, he wound down research at the mine soon afterwards, and continued his neutrino studies at US sites.

Cole doesn't dig deeply enough into these issues, so it is hard understand exactly how they were resolved. He also scatters descriptions of the relevant physics rather unsatisfactorily across chapters.

Nonetheless, *Chasing the Ghost* nicely describes how successful the derring-do attitude of individual researchers can be. Reines variously comes across as endearing, admirable and irritating. He could alarm his team by tugging on cables to test electronics as he whistled his way through a lab, and was quick to over-interpret results. But he was a hands-off, respectful lab chief who addressed his team with old-fashioned formality as 'Mr' (they seem to have all been men) — even as he ignored their rights to holidays. Those interviewed all tell how they fell under his spell, and worked hard to please him. One anecdote has him on one knee, singing an aria at a party. Unlike his neutrinos, it seems, Reines was always an unmistakable presence.

Nature **593**, 334-335 (2021)

doi: <https://doi.org/10.1038/d41586-021-01318-y>

## Jobs from Nature Careers

- - - [All jobs](#)
    - 
    - [Post-Doc - Numerical and microfabrication development of 4D metamaterials for mechanical, acoustic and photonics applications](#)
      - [Université Bourgogne Franche-Comté \(UBFC\)](#)
      - [BESANCON, France](#)
      - [JOB POST](#)

[The University of British Columbia \(UBC\)](#)

[Vancouver, Canada](#)

[JOB POST](#)

- [Postdoctoral Fellow | Zandstra Stem Cell Bioengineering Lab](#)

[The University of British Columbia \(UBC\)](#)

[Vancouver, Canada](#)

[JOB POST](#)

- [Masterthesis / internship \(m/f/x\)](#)

[Helmholtz Centre for Environmental Research \(UFZ\)](#)

[Leipzig, Germany](#)

[JOB POST](#)

[Download PDF](#)

[Download PDF](#)

## Related Articles

-  [Frederick Reines \(1918-98\)](#)

-  **Neutrinos could shed light on why the Universe has so much more matter than antimatter**
-  **Japan will build the world's largest neutrino detector**

---

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-021-01318-y>.

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

# Opinion

- [\*\*C. Austen Angell \(1933–2021\)\*\*](#) [ 14 May 2021]  
Obituary • Visionary explorer of glasses and the limits of the liquid state.
- [\*\*Craft an African American Graves Protection and Repatriation Act\*\*](#) [ 19 May 2021]  
Comment • Universities and museums must catalogue the remains of Black Americans in their collections, and pause research pending consultation with descendant communities.
- [\*\*UNESCO embraces open science to shape society's future\*\*](#)  
[ 18 May 2021]  
Correspondence •
- [\*\*Contrails: tweaking flight altitude could be a climate win\*\*](#)  
[ 18 May 2021]  
Correspondence •
- [\*\*Statistics: unify ecosystems valuation\*\*](#) [ 18 May 2021]  
Correspondence •
- [\*\*Virus variants: GISAID policies incentivize surveillance in global south\*\*](#) [ 18 May 2021]  
Correspondence •

## OBITUARY

14 May 2021

# C. Austen Angell (1933–2021)

Visionary explorer of glasses and the limits of the liquid state.

- [Pablo G. Debenedetti](#) <sup>0</sup>,
- [Peter H. Poole](#) <sup>1</sup>,
- [Srikanth Sastry](#) <sup>2</sup> &
- [Francesco Sciortino](#) <sup>3</sup>

### 1. [Pablo G. Debenedetti](#)

1. Pablo G. Debenedetti is dean for research at Princeton University, New Jersey, USA.

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

### 2. Peter H. Poole

1. Peter H. Poole is a professor in the physics department at St Francis Xavier University in Nova Scotia, Canada.

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

### 3. Srikanth Sastry

1. Srikanth Sastry is a professor in the theoretical sciences unit at the Jawaharlal Nehru Centre for Advanced Scientific Research in Bengaluru, India.

[View author publications](#)

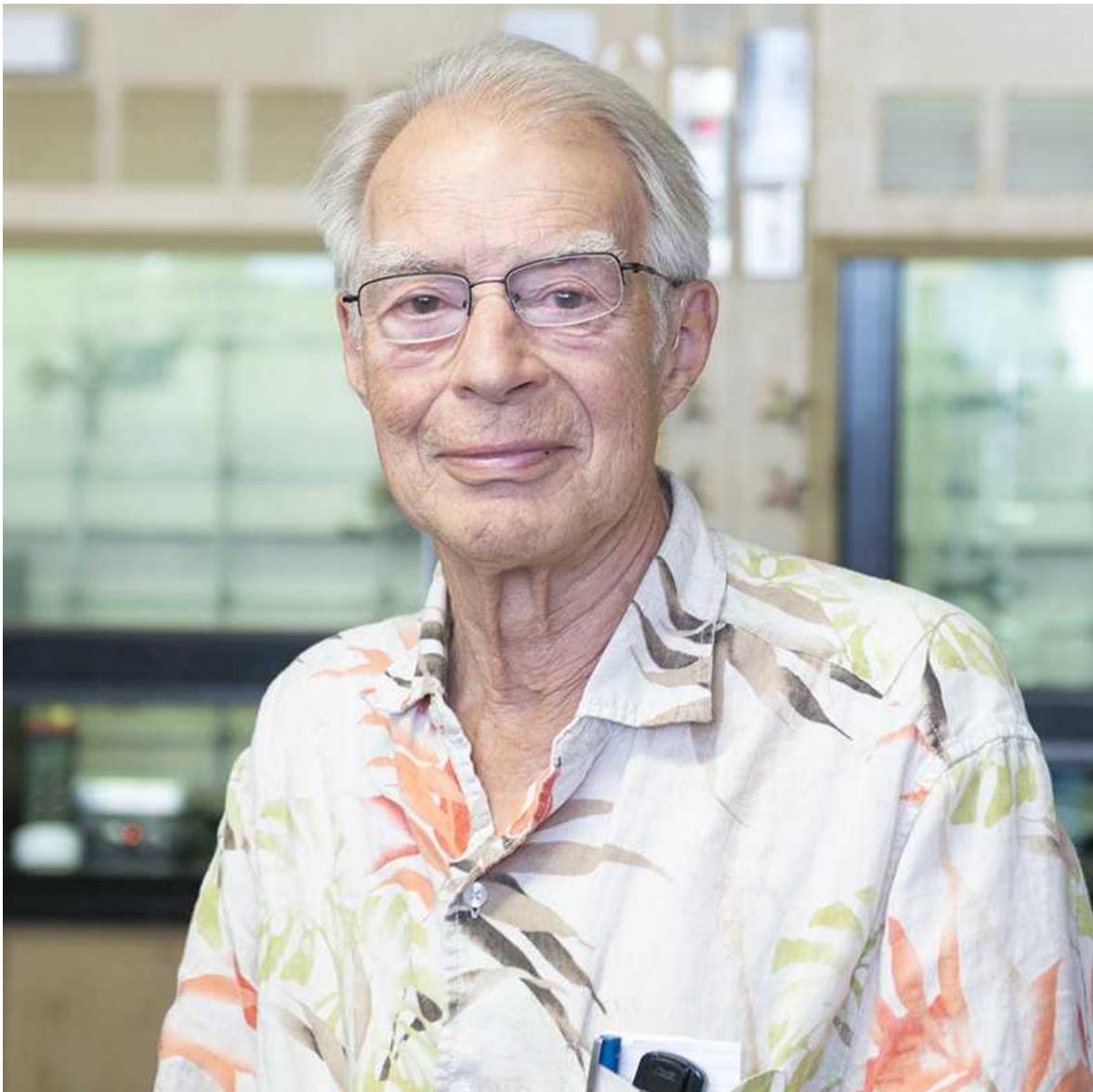
You can also search for this author in [PubMed](#) [Google Scholar](#)

4. Francesco Sciortino

1. Francesco Sciortino is professor of condensed-matter physics at Sapienza University of Rome, Italy.

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

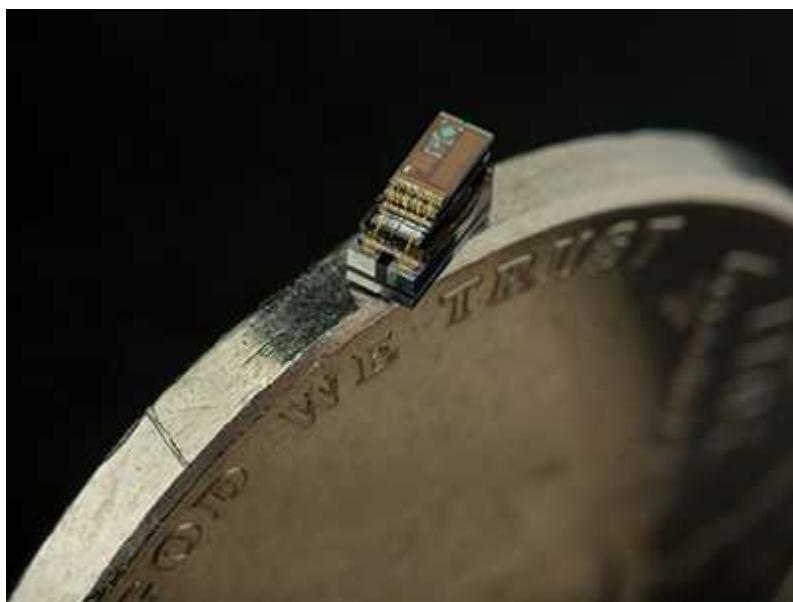


Credit: Mary Zhu

Austen Angell was one of the most versatile physical chemists of his generation. He made pioneering experimental discoveries on glasses and liquids, forms of matter that continue to challenge researchers in many disciplines. His work shaped the ideas that chemists, physicists and engineers use to investigate these materials, both in fundamental research and in applications such as fuel-cell and battery technology. He has died aged 87.

Liquids form amorphous solids, or glasses, when cooled fast enough that they do not crystallize. Being both dense and disordered, liquids and glasses cannot be explained using theories of crystals, which are dense and ordered, or gases, which are disordered but not dense. As a result, long-standing questions remain hotly debated — about the nature of the transition from liquid to glass, and how that most ubiquitous of liquids, water, behaves at low temperatures. For more than five decades, Angell led the search for a unified understanding of these problems, and he contributed many discoveries and concepts that now define the field.

Angell was born and raised in Canberra. His father, a plant pathologist and avid hobbyist, stimulated his interest in hot liquids by casting aluminium objects from aircraft scrap, melted in the dining-room fireplace. Angell studied chemical metallurgy at the University of Melbourne, Australia, then worked on molten salts with John Bockris at the University of Pennsylvania in Philadelphia. He completed his PhD with John Tomlinson at Imperial College London in 1961. Returning to the University of Melbourne, he was fascinated by fibres that he pulled from a melt of potassium nitrate and calcium nitrate, simple inorganic molecules that readily form glasses. From then on, he said, “I was hooked on glass-formers”.



[Tiny robots and sensors need tiny batteries — here's how to do it](#)

Angell moved to Purdue University in West Lafayette, Indiana, in 1966. He started work on aqueous solutions, many of which easily form glasses. He also considered pure water, in which the transition to a glass is notoriously difficult to achieve. Water has many unusual properties, such as reaching its maximum density at 4 °C. With a succession of talented students and postdocs, including Robin Speedy, Angell showed that water becomes even more unusual as it is supercooled — kept liquid below its freezing point. As the temperature falls, key properties such as specific heat and compressibility rise ever faster. This suggested the presence of a thermodynamic limit beyond which supercooled water cannot remain liquid, an idea now known as the Speedy–Angell conjecture. The work triggered intense interest in supercooled and glassy water that continues today, including the possibility of a transition from one liquid phase to another.

Angell had an unusual ability to perceive patterns that connected widely differing systems. In what is now called an Angell plot, he charted how the viscosity of liquids changed with temperature for different glass-transition temperatures, allowing the behaviours of glass-forming liquids to be easily compared. This revealed that liquids form a continuous spectrum, from “strong liquids” that approach the glass transition with a gradual variation in viscosity to “fragile liquids”, which display precipitous and accelerating viscosity changes as the transition is approached. The compelling simplicity of Angell’s classification has shaped research on glass formers, from practical materials engineering to the theoretical modelling of glasses by physicists.

Angell’s interest in the diffusion and conduction of ions in liquids and glasses was a constant thread. Over time, he applied it to systems such as molten salts, ionic liquids, ion-conducting glasses and battery electrolytes. With an eye to both the fundamental and the practical, his work established theoretical upper limits on the conductivity of solid electrolytes, such as ion-conducting glasses, that continue to guide the development of modern battery technology.



### Ten years left to redesign lithium-ion batteries

Although primarily an experimentalist, Angell was quick to realize the potential of computers. With colleagues in the 1970s, he conducted some of the first simulations of molten silica and silicates, blazing a trail that computational geologists soon followed. From then on, Angell regularly collaborated in computational studies of glassy behaviour and of water, in search of the mechanisms underlying experimental results.

After Angell moved to Arizona State University in Tempe, in 1989, he learnt from local geochemists about micrometre-sized fluid inclusions in quartz. He and his wife and colleague, Jenny Green, exploited these to study water under kilobars of negative pressure (tension rather than compression).

Angell exported ideas developed in the study of glasses to protein folding, cryopreservation and plastic crystals. He proposed that some liquids, including water, undergo a fragile-to-strong transition on cooling, and integrated this with the related phenomenon of glasses that exist in more than one amorphous state. Active until the end, Angell made use of these ideas to understand devices that depend on a change in phase for memory storage.

An irrepressible explorer, he loved to travel, to the benefit of many conferences. Angell was a great connector and instigator, always linking disparate ideas. At conferences, his post-talk questions invariably provided

fresh perspective. He was not afraid to speculate. Quipped Speedy, his long-time colleague: “Angell rushes in where fools fear to tread.”

Angell was gregarious and generous. Combined with his charisma, humility and humour, his love of discovery was infectious. He made a point of connecting with young researchers and promoting their work. His thoughtful advice influenced many careers. To his many students, postdocs and colleagues, Angell was an icon, a source of inspiration and a true friend.

Nature **593**, 336 (2021)

doi: <https://doi.org/10.1038/d41586-021-01319-x>

## Jobs from Nature Careers

- - - [All jobs](#)
    - - [Post-Doc - Numerical and microfabrication development of 4D metamaterials for mechanical, acoustic and photonics applications](#)  
[Université Bourgogne Franche-Comté \(UBFC\)](#)  
[BESANCON, France](#)  
[JOB POST](#)
      - [Postdoctoral Position - Ecological Modeler](#)  
[The University of British Columbia \(UBC\)](#)  
[Vancouver, Canada](#)  
[JOB POST](#)

- [Postdoctoral Fellow | Zandstra Stem Cell Bioengineering Lab](#)

[The University of British Columbia \(UBC\)](#)

[Vancouver, Canada](#)

[JOB POST](#)

- [Personalreferent\\*in \(d/m/w\), Gruppe Personaladministration](#)

[Helmholtz Centre for Heavy Ion Research GmbH \(GSI\)](#)

[Darmstadt, Germany](#)

[JOB POST](#)

[Download PDF](#)

[Download PDF](#)

## Related Articles

-  [Ten years left to redesign lithium-ion batteries](#)
-  [COVID-19 disruptions to tech-metals supply are a wake-up call](#)



- **Tiny robots and sensors need tiny batteries**  
**— here's how to do it**

---

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-021-01319-x>

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

COMMENT

19 May 2021

# Craft an African American Graves Protection and Repatriation Act

Universities and museums must catalogue the remains of Black Americans in their collections, and pause research pending consultation with descendant communities.

- [Justin Dunnivant](#) <sup>0</sup>,
- [Delande Justinvil](#) <sup>1</sup> &
- [Chip Colwell](#) <sup>2</sup>

## 1. [Justin Dunnivant](#)

1. Justin Dunnivant is a postdoctoral fellow in historical archaeology in the Department of Anthropology, Vanderbilt University, Nashville, Tennessee, USA.

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

## 2. [Delande Justinvil](#)

1. Delande Justinvil is a doctoral student in the Department of Anthropology, American University, Washington DC; and an incoming predoctoral fellow at the National Museum of Natural History, Washington DC, USA.

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

### 3. [Chip Colwell](#)

1. Chip Colwell is an anthropologist and editor-in-chief of *SAPIENS* magazine, New York City, New York, USA.

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)



At Sweet Briar Plantation Burial Grounds in Virginia, Bethany Pace works to protect the graves of her ancestors, many of whom were enslaved. Credit: Michael S. Williamson/*The Washington Post* via Getty

Last month, shocking news reports revealed that what are thought to be the skeletal remains of Tree and Delisha Africa, two Black girls killed in a US police bombing in 1985, might have been studied for years by researchers at two US universities, without their families' permission. The finding, involving the University of Pennsylvania in Philadelphia and Princeton University in New Jersey, is just the latest in a series of discoveries in

university collections related to the mistreatment of African American human remains.

A week earlier, the University of Pennsylvania announced that it would rebury the remains of more than 50 enslaved people held in its anthropology museum. In January, Harvard University in Cambridge, Massachusetts, announced that it was creating a committee to consider policies around its museum collections after the discovery of the remains of 15 people who were enslaved. In 2017, the University of Virginia in Charlottesville acknowledged that it had lost track of grave-robbed remains from African American cemeteries, which the medical school had once used for anatomical study<sup>1</sup>.

The call for institutional accountability over African American remains in academic collections comes at a time when the US Congress will soon convene hearings on the African American Burial Grounds Network Act. This bill would survey and offer recommendations for the protection of African American burial grounds. It is a good first step.

Although this could become one of the most significant pieces of legislation in the fight to safeguard Black heritage, the United States needs much stronger laws to respectfully care for the graves, osteological remains and genetic material of deceased African Americans. There are thousands of remains in unmarked burial grounds and institutional collections around the country, which are at risk of loss, negligence and destruction (see ‘Defend the dead’).

## DEFEND THE DEAD

Across the United States, the remains of thousands of Black people have been discovered in unmarked graves or unofficial cemeteries. These sites are often razed by development. Only a few have been granted protection. There could be many more.



**3**  
Civil War Union soldiers. Found 2015. Rosedawn Cemetery, Pueblo, Colorado

**65**  
Convict leasing labourers. Buried 1878–1911, found 2018. Fort Bend Independent School District, Sugar Land, Texas

**173**  
Enslaved people and their descendants. Buried 1852 onwards, found 2015. Oakwood Cemetery, Tyler, Texas

**90+**  
Enslaved people. Found 2019. Christian Life Cathedral, Fayetteville, Arkansas

**1,000+**  
Enslaved people. Found 2013. Ascension Parish Plantations, Lemariville, Louisiana

**6**  
**119**  
Enslaved people. Buried 1848 onwards, found 2019. Red Banks Cemetery, Red Banks, Mississippi

**7**  
**13**  
Victims of the Chicago Race Riot. Buried 1919, found 2019. Lincoln Cemetery, Blue Island, Illinois

**8**  
**12+**  
First African American settlers. Buried early 19th century, found 2019. Butternut Ridge Cemetery, North Olmsted, Ohio

**9**  
**100+**  
Enslaved people. Found 2015. District Hill Cemetery, Chickamauga, Georgia

**10**  
**872**  
Buried 1866 onwards, found 2016. Oakland Cemetery, Atlanta, Georgia

**11**  
**1,146**  
Enslaved and segregated people. Buried 1870s–1950s. Alta Vista Cemetery, Gainesville, Georgia

**12**  
**100+**  
Enslaved or formerly enslaved people. Found 2015. Old Athens Cemetery, University of Georgia, Athens, Georgia

**13**  
**363+**  
Formerly enslaved people. Buried 1870s–1900s, found 2006. Hunter Army Airfield, Savannah, Georgia

**14**  
**200–250**  
Enslaved and convict leasing labourers. Buried 1850s. Woodland Cemetery, Clemson, South Carolina

**15**  
**900+**  
Enslaved, formerly enslaved and Black parishioners. Buried 1800s onwards, found 2013. Soapstone Baptist Church, Pickens, South Carolina

**16**  
**-100**  
Enslaved people. Found 2020. Eddy Lake Cemetery, Bucksport, South Carolina

**17**  
**40+**  
Enslaved people. Buried 1830s onwards, found 2019. Capital City Country Club, Tallahassee, Florida

**18**  
**44**  
Segregated people. Found 2020. Pinellas County School District, Clearwater, Florida

**19**  
**120+**  
Segregated people. Found 2019. Robles Park Village, Tampa, Florida

**20**  
**145+**  
Paupers' burial ground. Buried 1950s, found 2019. King High School, Tampa, Florida

**21**  
**674**  
Buried 1928, found 1950s. Hurricane of 1928 Mass Burial Site, West Palm Beach, Florida

**22**  
**165**  
Enslaved people. Found 2016. Historic Black Cemetery, Boone, North Carolina

**23**  
**19+**  
Enslaved people. Found 2017. Mallard Creek Presbyterian Church, Charlotte, North Carolina

**24**  
**200+**  
Enslaved people. Buried after 1959, found 2018. St Philips Moravian Graveyard, Winston-Salem, North Carolina

**25**  
**475**  
Enslaved people. Buried after 1859, found 2015. University of North Carolina at Chapel Hill, North Carolina

**26**  
**32**  
Enslaved people. Found 2005. Avoca Museum, Altavista, Virginia

**27**  
**67**  
Enslaved people. Buried 19th century, found 2012. University of Virginia, Charlottesville, Virginia

**28**  
**200+**  
Segregated veterans and families. Found 2020. Ellsworth Cemetery, Westminster, Maryland

**29**  
**5,000+**  
Enslaved, formerly enslaved and veterans. Buried 1852 onwards, found 2018. Laurel Cemetery, Baltimore, Maryland

**30**  
Enslaved people. Buried 19th century, found 2019. Former Belvoir Plantation, Crownsville, Maryland

**31**  
Enslaved people. Buried 19th century, found 2018. University of Pennsylvania, Philadelphia, Pennsylvania

**32**  
**673**  
Nation's first free Black settlers. Buried from 1800s, found 2018. Rossville A.M.E. Zion Church Cemetery, Staten Island, New York

**33**  
**15,000+**  
Mostly enslaved people. Buried 17th–18th centuries, found 1897. African Burial Ground National Monument, New York City, New York

Source: Data and map: Hanna Kim & the Convict Leasing and Labor Project

We propose in addition the creation of an African American Graves Protection and Repatriation Act (AAGPRA), modelled on existing federal legislation for Native American remains and cultural items. Such a law would protect graves and provide guidance on the care and repatriation of human remains in scientific collections. It could do so in a manner that also addresses a growing interest in genetic samples, both for genealogical testing services and for medical and historical research.

The practices of academic institutions should match their bold statements against anti-Black racism. The scientific community must embrace policies that catalyse new collaborations to honour community needs and desires. Greater accountability among researchers, universities and museums is long past due. Here's why this should be done, and how.

## No protection

Because of historical oppression in US society, the graves of enslaved people and their descendants were often unmarked, placed in unofficial cemeteries and razed by development projects. Across the United States, from the late nineteenth century and continuing today, Black cemeteries were landfillled to make space for buildings, levelled for the creation of parks, and either destroyed or put at risk by suburban developments, roads, infrastructure and housing<sup>2</sup>. Many burial sites held hundreds of people, a few held several thousand<sup>3</sup>.

Through the years, Black descendants and concerned residents formed associations and coalitions in defence of these burial sites. African American cemeteries have struggled to acquire state and federal recognition<sup>4</sup>. For example, to protect cemeteries, the US National Historic Preservation Act of 1966 requires evidence that they achieved historic significance, are associated with historic events, have the potential to yield historic information, or contain the graves of people of “transcendent importance”. For marginalized, poorly documented graveyards, this is difficult — or impossible. Moreover, significant individuals in disenfranchised minority communities are often not recognized by those in power.

Since the early 1990s, high-profile archaeological projects have led national calls for the preservation of such sites. These include the New York African Burial Ground<sup>5</sup>, Freedman's Cemetery<sup>6</sup> in Dallas, Texas, and the First African Baptist Church graveyard<sup>7</sup> in Philadelphia. At the latter, for instance, researchers worked with community leaders to protect graves while studying the historical experiences of urban enslaved and free African Americans from the eighteenth century onwards. Last year, scholars in Oklahoma involved in the discovery of a probable mass grave of Black Americans — who could well be among hundreds killed by a white mob in the 1921 Tulsa Race Massacre — joined in the call.

There are still no federal protections specifically for historic Black cemeteries.



Laurel Hill Cemetery in Baltimore, Maryland, was demolished in 1958 to build a market. Credit: Afro American Newspapers/Gado/Getty

Today, the remains of at least 2,000 African Americans — possibly many more — are in museums, medical collections and universities around the United States. Black burial grounds were plundered for research until the early twentieth century. For instance, construction crews discovered thousands of remains at the Medical College of Georgia in Augusta in 1989, most of which had been stolen from a graveyard for Augusta's poor and Black citizens between 1835 and 1913. In other cases, the remains of Black people did not even reach the grave before being taken into collections. For example, the body of Nat Turner — the freedom fighter who was hanged and skinned in 1831 for leading a rebellion — is thought to have entered the ‘cadaver trade’, which supplied US anatomy classrooms<sup>8</sup>.

In the 1980s, it became possible to extract DNA from osteological samples for genealogical research. When genomic data are coupled with biochemical and archaeological studies, they can paint a complex portrait of demography, disease, geographical origin and more. These advances, coupled with interest in African American ancestry, spurred public and private sequencing of DNA from African American human remains<sup>9</sup>.

But the acquisition, storage and use of this genetic material cannot be undertaken lightly. The promise of expanding knowledge must be considered alongside the perils of medical exploitation and biocolonialism<sup>10</sup>. Recall the globalization of Henrietta Lacks's tumour cells in lab research, or the 1840s gynaecological experimentation on Anarcha Westcott, an enslaved woman. Neither of these African Americans consented to their participation in research; both were reduced to scientific data, rather than being treated as humans with rights and relatives<sup>11,12</sup>. Modern extractive practices risk recapitulating the scientific racism that dehumanized Black lives.

This is a civil- and human-rights issue. The remains of African American people, as with those of Native Americans, have not received the same protections as the bodies of white citizens.

## Legislative fix

In December last year, the US Senate unanimously passed the African American Burial Grounds Study Act. But the House of Representatives

adjourned before the legislation could be considered. In Congress's 2021 term, the bill — now provisionally titled the African American Burial Grounds Network Act — is anticipated to be reintroduced with the support of an alliance of dozens of scientific, community and national organizations. Research champions for the bill include the US National Trust for Historic Preservation, the Society for Historical Archaeology and the Society of Black Archaeologists.

If successful, the legislation would allocate funding for the creation of a voluntary network to build a national database of African American burial grounds. It would also consider best practices for commemoration and preservation.

This is a crucial step to redress the centuries of desecration perpetrated against African American cemeteries and skeletal remains.

More is needed. The legislation does not protect cemeteries on federal lands or mandate consultations with descendants. Nor does it address the thousands of human remains in museums and research institutions. A better model is the Native American Graves Protection and Repatriation Act (NAGPRA) of 1990. This US federal law gives certain rights to descendants and tribes over ancestral human remains, in the ground and in museums.



Howard University students Miesha Hegwood and Ike Mesumbe surveyed a former African American cemetery at Walter Pierce Park in Washington DC in 2006. Credit: Robert A. Reeder/The Washington Post via Getty

We call for the creation of an analogous African American Graves Protection and Repatriation Act (AAGPRA). It could emulate such provisions as a grant programme, prohibitions against the sale and transfer of human remains, and a review committee to oversee the law and adjudicate disputes.

One principle would be the inventory process. A network of archaeologists, city planners and communities could together identify burial grounds and plan for their stewardship. This could build on previous efforts such as the National Burial Database of Enslaved Americans and the Convict Leasing and Labor Project. For federal agencies and federally funded museums, AAGPRA would require a full public inventory of their human-remains collections. As under NAGPRA, these inventories could be published in the Federal Register and listed on a National Park Service website.

A second principle, similar to NAGPRA, would be consultation with descendants and community leaders. This would require researchers and institutions to actively prioritize inclusive decision-making processes. An AAGPRA could build on protocols developed by leading scholar-advocates that foster genuine engagement.

For example, through meaningful collaboration between archaeologists, artists, community leaders and government officials, the New York African Burial Ground Project was driven by a ‘clientage model’. This replaced the sort of tokenistic engagement that project director Michael Blakey has described as “hearing Blacks but not listening, looking at Blacks but not seeing”<sup>13</sup>. Clientage led to discoveries such as how the changing patterns of gruelling forced labour bent the bodies of enslaved men and then women. And it led to memorialization — the site became a US national monument with a visitor centre<sup>13</sup>.

## Challenges ahead

Determining who exactly should be consulted presents several challenges for African Americans. Under NAGPRA, the primary consulting parties are sovereign tribal nations, and cultural relationships between claimants and ancestral remains are determined through analysis of ten lines of evidence. Centuries of displacement and sparse genealogical records for African Americans can mean that it is difficult to link a set of human remains to specific Black descendants<sup>14</sup>.

Those working with African American remains can draw from the National Trust for Historic Preservation’s 2018 publication, *Engaging Descendant Communities* (see [go.nature.com/3g76zt4](https://go.nature.com/3g76zt4)). This provides guidelines for multidisciplinary research that is collaborative, transparent, accountable and accessible. It defines descendants both in genealogical terms and more inclusively, to welcome input from African Americans whose ancestors had a shared historical experience. Community organizations, churches and national groups should be empowered to guide decisions on potential solutions ranging from retaining collections to reburial.

Genetics could help to establish biological relatedness between individuals in museums and current kin. All involved must acknowledge that it is a destructive method — requiring bone to be sampled and ground up. And results can be disappointing — genetic reference libraries generally provide only a snapshot sample of populations, and predominantly hold data on people of European ancestry.

Delineating how the genetic data of living and deceased African Americans could and should be used is essential. It is key that this powerful tool does not replicate the very practices it is intended to redress<sup>[15](#)</sup>. As we have learnt from Indigenous scientists, conversations of data governance must unfold for Black communities to reduce the risk of yet another form of biological data being further out of their control<sup>[16](#)</sup>.



Pall-bearers accompanied the remains of 419 free and enslaved African Americans who were reburied in a 2003 ceremony at the African Burial Ground Memorial Site in New York City. Credit: Stan Honda/AFP via Getty

Given that NAGPRA has been controversial — often seemingly pitting scientific goals against Indigenous sovereignty — one could reasonably ask why this law for Native Americans should be a model for African Americans. In our view, NAGPRA is not anti-science, it is pro-human<sup>17</sup>. Like the human-subjects regulations that established necessary oversight after the betrayal of the [Tuskegee syphilis experiment in Alabama](#) (which left hundreds of African American men with the disease untreated), NAGPRA does not prohibit research. Rather, it requires scientists to respect the dignity of all people and seek the consent of descendants. We contend that this should be considered a necessary component of good science.

The collaborative ethic that emerged among Native Americans and museum professionals following NAGPRA, despite the act's flaws, offers hope for how an African American analogue might bring together diverse stakeholders equitably. Exemplars of co-production since 1990 have resulted in discoveries — such as information about the ancestry, descendants and seafood diet of a man whose 10,300-year-old remains were excavated in Alaska — and celebrations, such as the revival of a lost Indigenous boat-building technique used by the Sugpiat people of south central Alaska.

In a similar vein, at Oakwood Cemetery in Austin, Texas, Black descendants are helping to guide work with archaeologists, geneticists and city officials to recover the identities and lived experiences of 36 marginalized people buried in the mid- to late nineteenth century. The burials were discovered under a chapel during its restoration, and the collaboration explores questions related to these people's diet, experiences of disease, stress and trauma, and potential genetic relatedness to living populations (see [go.nature.com/3uedoyu](http://go.nature.com/3uedoyu)).

## Four steps

To develop an AAGPRA, we suggest these four priorities for the next two years.

**Pass the African American Burial Grounds Network Act.** To protect African American burial sites and remains, we must know where they are,

and understand their condition and any potential threats to them. Therefore, the passage of the act is a logical first step.

**Catalogue existing osteological collections.** We think that museums, universities and other institutions should voluntarily survey their existing osteological collections and publish public summaries. They should reassess their collection protocols in consultation with descendants. Where possible, they should undertake genealogical and historical investigations to identify potential descendants. This work will certainly identify further African American remains and establish best practices that could be used as the foundation for a required inventory process under AAGPRA.

**Pause the unethical study and use of Black remains in collections.** In instances where ethical and legal standards have not been determined by stakeholders, there should be a moratorium on destructive research or teaching that uses African American remains. Efforts should first exhaust all non-invasive methods to find known descendants or representative descendant groups, so that they can be consulted closely on giving consent for inventories or repatriation.

**Amend and extend other federal legislation.** Without risking a delay in passing the African American Burial Grounds Network Act, some key elements of an AAGPRA could be included now. But we need full and exhaustive solutions to this crisis. Scientists and community leaders must work together to apply the lessons learnt from these efforts for inventorying, conservation, consultation, research and repatriation when necessary. Federal legislation must also consider the ethics around genetic testing and genealogical research<sup>18</sup>. These efforts will require coordination and mutual support among all constituencies.

## Time to act

Recommendations of this kind have been discussed in the three decades since Native American remains were legally, at least, afforded the respect we're calling for here. Researchers affiliated with the New York African Burial Ground Project and other archaeological sites have pioneered

laudable case studies that demonstrate how scientists and communities can work together<sup>[19,20](#)</sup>.

What is new is the US political momentum. A new president and Congress seem committed to combating anti-Black racism. There has been nationwide protest over police violence, worldwide protest over systemic racism, renewed public outrage over Black people's remains being held in university collections, and this year is the centennial of the Tulsa Race Massacre. All this, we hope, will foster a united effort to bring dignity to the ancestors of African Americans (see [Editorial](#) and [Nature https://doi.org/dx7r; 2020](https://doi.org/dx7r; 2020)).

The past year has seen numerous leading academic and scientific institutions make statements about the need to end systemic racism and advance social justice. The same institutions can act on their stated commitments by developing, or reassessing, their protocols for acquisition, conservation and research, in alignment with an ethic of care. They can use their power to leverage support for the African American Burial Grounds Network Act, the foundation for a long-term vision of providing civil and human rights to Black Americans in death as well as in life.

Nature **593**, 337-340 (2021)

doi: <https://doi.org/10.1038/d41586-021-01320-4>

## References

1. 1.

Gates, E. 'Theatre of the Macabre' (*University of Virginia Magazine* 106, Spring 2017).

2. 2.

Jones, A. 'Stop Destroying African American Cemeteries' *SAPIENS* (23 February 2021).

3. 3.

Harrington, S. P. M. *Archaeology* **46**, 28–39 (1993).

4. 4.

Rainville, L. J. *Field Archaeol.* **34**, 196–206 (2009).

5. 5.

Blakey, M. L. *Transform. Anthropol.* **7**, 53–58 (1998).

6. 6.

Davidson, J. M. *Int. J. Hist. Archaeol.* **11**, 193–220 (2007).

7. 7.

Rankin-Hill, L. M. In *New Directions in Biocultural Anthropology* (eds Zuckerman, M. K. & Martin, D. L.) 133–156 (Wiley, 2016).

8. 8.

Berry, D. R. *The Price for Their Pound of Flesh* (Beacon, 2017).

9. 9.

Nelson, A. *The Social Life of DNA* (Beacon, 2016).

10. 10.

Clinton, C. K. & Jackson, F. L. C. *Am. J. Phys. Anthropol.*  
<https://doi.org/10.1002/ajpa.24171> (2020).

11. 11.

Skloot, R. *The Immortal Life of Henrietta Lacks* (Crown, 2010).

12. 12.

Owens, D. C. *Medical Bondage* (Univ. Georgia Press, 2017).

13. 13.

Blakey, M. L. *Curr. Anthropol.* **61**, S183–S197 (2020).

14. 14.

Watkins, R. J. In *Identified Skeletal Collections* (eds Henderson, C. Y. & Cardoso, F. A.) 169–186 (Archaeopress Archaeology, 2018).

15. 15.

Roberts, D. *Fatal Invention* (New Press, 2012).

16. 16.

TallBear, K. *Native American DNA* (Univ. Minnesota Press, 2013).

17. 17.

Colwell, C. *Plundered Skulls and Stolen Spirits* (Univ. Chicago Press, 2017).

18. 18.

Fox, K. & Hawks, J. *Nature* **572**, 581–583 (2019).

19. 19.

Blakey, M. L. In *Evaluating Multiple Narratives* (eds Habu, J., Fawcett, C. & Matsunaga, J. M.) 17–28 (Springer, 2008).

20. 20.

Watkins, R. J. *Hist. Archaeol.* **54**, 17–33 (2020).

## Jobs from Nature Careers

- -

- [All jobs](#)
  - 
  - [Post-Doc - Numerical and microfabrication development of 4D metamaterials for mechanical, acoustic and photonics applications](#)

[Université Bourgogne Franche-Comté \(UBFC\)](#)

[BESANCON, France](#)

[JOB POST](#)

- [Postdoctoral Position - Ecological Modeler](#)

[The University of British Columbia \(UBC\)](#)

[Vancouver, Canada](#)

[JOB POST](#)

- [Postdoctoral Fellow | Zandstra Stem Cell Bioengineering Lab](#)

[The University of British Columbia \(UBC\)](#)

[Vancouver, Canada](#)

[JOB POST](#)

- [Personalreferent\\*in \(d/m/w\), Gruppe Personaladministration](#)

[Helmholtz Centre for Heavy Ion Research GmbH \(GSI\)](#)

[Darmstadt, Germany](#)

## JOB POST

[Download PDF](#)

[Download PDF](#)

## Related Articles



- [Divided by DNA: The uneasy relationship between archaeology and ancient genomics](#)



- [What Black scientists want from colleagues and their institutions](#)



- [Weapons for when bigotry claims science as its ally](#)



- [Ethics of exhibition](#)



- [Genomics: DNA and diasporas](#)

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

## CORRESPONDENCE

18 May 2021

# UNESCO embraces open science to shape society's future

- [Audrey Azoulay](#) 0

1. [Audrey Azoulay](#)

1. UNESCO, Paris, France.

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

After a transparent and inclusive consultation process, the 193 member states of the United Nations cultural organization UNESCO started negotiating the final text of its ‘Recommendation on Open Science’ this month. This global standard-setting instrument, which seeks to close the gaps in science, technology and innovation within and between nations, will be adopted at the 41st UNESCO General Conference in November. The scientific community’s input continues to be crucial (see [go.nature.com/3oatqrk](http://go.nature.com/3oatqrk)).

The COVID-19 pandemic has underscored the unacceptable inequalities affecting access to scientific progress, knowledge and education. It has forced us to rethink how we produce, share and communicate scientific information. Universal access to science must be more inclusive, more collaborative and more interdisciplinary.

Although key to prosperity, peace and a healthy planet — all priorities for the UN’s 2030 Agenda for Sustainable Development — open science still has no common definition or international policy framework. UNESCO’s

recommendation aims to build a shared set of principles based on quality, integrity, collective benefit and diversity. It will support policymaking and investment in infrastructure, services, educational resources and capacity-building.

We must harness open science's potential to meet societal needs and shape the future, backed by civil society, policymakers, young people and the scientific community.

Nature **593**, 341 (2021)

doi: <https://doi.org/10.1038/d41586-021-01338-8>

## Jobs from Nature Careers

- - - [All jobs](#)
    - 
    - [Post-Doc - Numerical and microfabrication development of 4D metamaterials for mechanical, acoustic and photonics applications](#)

[Université Bourgogne Franche-Comté \(UBFC\)](#)

[BESANCON, France](#)

[JOB POST](#)

- [Postdoctoral Position - Ecological Modeler](#)

[The University of British Columbia \(UBC\)](#)

[Vancouver, Canada](#)

[JOB POST](#)

- **Postdoctoral Fellow | Zandstra Stem Cell Bioengineering Lab**

The University of British Columbia (UBC)

Vancouver, Canada

JOB POST

- **Personalreferent\*in (d/m/w), Gruppe Personaladministration**

Helmholtz Centre for Heavy Ion Research GmbH (GSI)

Darmstadt, Germany

JOB POST

Download PDF

Download PDF

## Related Articles

- **See more letters to the editor**

---

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-021-01338-8>

## CORRESPONDENCE

18 May 2021

# Contrails: tweaking flight altitude could be a climate win

- [Ken Caldeira](#) <sup>0</sup> &
- [Ian McKay](#).<sup>1</sup>

1. [Ken Caldeira](#)

1. Carnegie Institution for Science, Stanford, California, USA.

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

2. Ian McKay

1. Gates Ventures, Kirkland, Washington, USA.

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

By our calculation, preventing most of the damaging impact of aircraft vapour trails (contrails) on climate would cost less than US\$1 billion per year and the net value of the benefit could be more than 1,000 times that. We know of no comparable climate investment with a similarly high likelihood of success.

Contrails account for roughly 2% of global warming because they cause cirrus clouds to trap outgoing infrared radiation. The flights responsible are typically within 1,000 metres of trajectories that are much less damaging to

the atmosphere, so planes routed to higher or lower altitudes would burn less than 0.1% more fuel ([R. Teoh et al. Environ. Sci. Technol. 54, 2941–2950; 2020](#)).

The warming effect of the extra emissions is likely to be more than offset by the reduction in vapour-trail formation (see [B. Kärcher Nature Commun. 9, 1824; 2018](#)). Furthermore, any drop in altitude would be insufficient to affect health costs from particulate matter and noise on the ground.

Addressing climate change could cost as much as US\$100 trillion ([D. P. van Vuuren et al. Nature Clim. Change 10, 329–334; 2020](#)). Abating even 1% of this cheaply could therefore represent a trillion-dollar economic opportunity.

Nature **593**, 341 (2021)

doi: <https://doi.org/10.1038/d41586-021-01339-7>

## Jobs from Nature Careers

- - - [All jobs](#)
    - 
    - [Post-Doc - Numerical and microfabrication development of 4D metamaterials for mechanical, acoustic and photonics applications](#)

[Université Bourgogne Franche-Comté \(UBFC\)](#)

[BESANCON, France](#)

[JOB POST](#)

- [Postdoctoral Position - Ecological Modeler](#)

The University of British Columbia (UBC)

Vancouver, Canada

JOB POST

- **Postdoctoral Fellow | Zandstra Stem Cell Bioengineering Lab**

The University of British Columbia (UBC)

Vancouver, Canada

JOB POST

- **Personalreferent\* in (d/m/w), Gruppe Personaladministration**

Helmholtz Centre for Heavy Ion Research GmbH (GSI)

Darmstadt, Germany

JOB POST

Download PDF

Download PDF

## Related Articles

- **See more letters to the editor**

---

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-021-01339-7>

CORRESPONDENCE  
18 May 2021

# Statistics: unify ecosystems valuation

- [Nils Brown](#)<sup>0</sup>,
- [Aldo Femia](#)<sup>1</sup>,
- [Dennis Fixler](#)<sup>2</sup>,
- [Ole Gravgård Pedersen](#)<sup>3</sup>,
- [Sven C. Kaumanns](#)<sup>4</sup>,
- [Gian Paolo Oneto](#)<sup>5</sup>,
- [Simon Schürz](#)<sup>6</sup>,
- [Francesco N. Tubiello](#)<sup>7</sup> &
- [Scott Wentland](#)<sup>8</sup>

1. Nils Brown

1. Statistics Sweden, Solna, Sweden.

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

2. Aldo Femia

1. Italian National Institute of Statistics, Rome, Italy.

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

3. [Dennis Fixler](#)

1. US Bureau of Economic Analysis, Suitland, Maryland, USA.

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

4. Ole Gravgård Pedersen

1. Statistics Denmark, Copenhagen, Denmark.

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

5. Sven C. Kaumanns

1. Federal Statistical Office of Germany, Bonn, Germany.

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

6. Gian Paolo Oneto

1. Italian National Institute of Statistics, Rome, Italy.

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

7. Simon Schürz

1. Federal Statistical Office of Germany, Bonn, Germany.

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

8. Francesco N. Tubiello

1. FAO Statistics Division, Rome, Italy.

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

9. Scott Wentland

1. US Bureau of Economic Analysis, Suitland, Maryland, USA.

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

Your Editorial decries as “unfortunate” the United Nations Statistical Commission’s decision to omit the valuation component from its international statistical standard for measuring ecosystems’ services and assets ([Nature 591, 178; 2021](#)). It calls out the United States and several European countries for objecting and delaying natural-capital accounting. We stand by the commission’s decision.

In our view, statisticians and national accountants worldwide should first agree on how to treat monetary values connected to ecosystem services.

International experience in valuing ecosystems at a national scale is sparse, and best practices have yet to emerge. The ecosystem component of the UN’s System of Environmental Economic Accounting needs broader consensus on a more rigorous statistical approach to valuation concepts and methods. Some are currently incompatible with the national accounts framework, limiting comparability with statistics such as gross domestic product.

We support further experimentation in ecosystems valuation that is grounded in the comparability and standardization required for robust statistical measurement. However, premature adoption of underdeveloped standards could damage progress in joint national and environmental economic accounting.

Nature **593**, 341 (2021)

doi: <https://doi.org/10.1038/d41586-021-01309-z>

## Jobs from Nature Careers

- - - [All jobs](#)
    - - [Post-Doc - Numerical and microfabrication development of 4D metamaterials for mechanical, acoustic and photonics applications](#)

[Université Bourgogne Franche-Comté \(UBFC\)](#)

[BESANCON, France](#)

[JOB POST](#)

  - [Postdoctoral Position - Ecological Modeler](#)

[The University of British Columbia \(UBC\)](#)

[Vancouver, Canada](#)

[JOB POST](#)

  - [Postdoctoral Fellow | Zandstra Stem Cell Bioengineering Lab](#)

[The University of British Columbia \(UBC\)](#)

[Vancouver, Canada](#)

[JOB POST](#)

▪ **Personalreferent\*in (d/m/w), Gruppe  
Personaladministration**

Helmholtz Centre for Heavy Ion Research GmbH (GSI)

Darmstadt, Germany

JOB POST

Download PDF

Download PDF

## Related Articles

- **See more letters to the editor**

---

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-021-01309-z>

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

## CORRESPONDENCE

18 May 2021

# Virus variants: GISAID policies incentivize surveillance in global south

- [Jeremy P. Kamil](#) 0

1. [Jeremy P. Kamil](#)

1. Louisiana State University Health Shreveport, Shreveport, Louisiana, USA; on behalf of 9 signatories.

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

When SARS-CoV-2 coronavirus genome sequences determined in low-income nations are rapidly shared through the independent Global Initiative on Sharing Avian Influenza Data (GISAID), attention and collaboration flow in from scientists at prestigious institutes (see, for example, [go.nature.com/3upedvz](https://go.nature.com/3upedvz)). This helps to highlight education and health-security inadequacies in less-wealthy countries.

Centralized workflows realize economies of scale, but adapt slowly to changes on the ground. In our view, the best way to fortify grass-roots participation and speed up data sharing would be for sequencing, analysis and discovery to occur in the communities from which samples are collected.

Policies benefiting those who share data incentivize local genomic surveillance. GISAID boosts sharing by negating researchers' reluctance to

rapidly deposit sequence data anonymously. Those calling for unrestricted data access ([Nature 590, 195–196; 2021](#)) propose conditions that have failed during public-health emergencies ([J. LoTempio et al. Sci. Diplom. 9, 47; 2020](#)).

Money is pouring in for genomic surveillance (see [go.nature.com/3up9g3g](#)). Greater community participation will underpin this investment by accelerating detection of new virus variants.

Nature **593**, 341 (2021)

doi: <https://doi.org/10.1038/d41586-021-01310-6>

A full list of co-signatories to this letter appears in Supplementary Information.

## Supplementary Information

1. [List of co-signatories](#)

## Competing Interests

V.S.C. is an adviser and co-founder of Microbial Genome Sequencing Center, LLC. J.P.K. is a consultant for BioNTech.

## [Jobs from Nature Careers](#)

- - - [All jobs](#)
    - 
    - [Post-Doc - Numerical and microfabrication development of 4D metamaterials for mechanical, acoustic and photonics applications](#)

Université Bourgogne Franche-Comté (UBFC)

BESANCON, France

JOB POST

▪ **Postdoctoral Position - Ecological Modeler**

The University of British Columbia (UBC)

Vancouver, Canada

JOB POST

▪ **Postdoctoral Fellow | Zandstra Stem Cell Bioengineering Lab**

The University of British Columbia (UBC)

Vancouver, Canada

JOB POST

▪ **Personalreferent\*in (d/m/w), Gruppe Personaladministration**

Helmholtz Centre for Heavy Ion Research GmbH (GSI)

Darmstadt, Germany

JOB POST

Download PDF

Download PDF

## Related Articles

- [\*\*See more letters to the editor\*\*](#)
- 

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-021-01310-6>

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

# Work

- **University of Cape Town's battle to tackle a racist legacy**

[ 18 May 2021]

Career Feature • Six years after a statue of Cecil Rhodes was toppled, students and staff at the South African university are still working to improve equity and representation.

- **Designing nano-sized chemotherapy** [ 17 May 2021]

Where I Work • Nanoparticles carrying chemotherapeutic drugs could help people with cancer escape some of the drugs' side effects, hopes Silvia Giordani.

---

CAREER FEATURE

18 May 2021

# University of Cape Town's battle to tackle a racist legacy

Six years after a statue of Cecil Rhodes was toppled, students and staff at the South African university are still working to improve equity and representation.

- [Linda Nordling](#)
  1. Linda Nordling  
[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

[Find a new job](#)



A statue of Cecil John Rhodes was removed from the campus of the University of Cape Town in 2015.Credit: Rodger Bosch/AFP/Getty

In 2015, a giant crane hoisted a 900-kilogram bronze statue of the imperialist Cecil John Rhodes from its plinth at the University of Cape Town (UCT) in South Africa. Rhodes — a nineteenth-century diamond magnate and a representative of the white-supremacist, colonial rule of southern Africa — had bequeathed the land on which the university now stands. The removal of his statue came two decades after South Africa's first-ever democratic elections and the end of apartheid.

When the crane did its work, Rhodes's likeness in front of UCT's main hall was reeking, both figuratively and literally. A month earlier, a student had upended a bucket of human excrement over the statue, lighting the fuse of what was to become known globally as the Rhodes Must Fall movement (see 'Of protest and potential'). Since the statue fell, UCT has played host to conversations about how to ensure that the institution — one of Africa's foremost — embraces inclusivity at its very core.

This includes challenging its traditions, which critics argue are rooted in colonial values and minimize African knowledge and experiences. These discussions have been difficult for both the academic establishment and those fighting against the status quo. Despite efforts to boost the number of Black researchers on campus, 40% of academic staff members are white South Africans, compared with 9% of the overall population in South Africa. Black South Africans make up more than 80% of the nation's population.

*Nature* asked four UCT students and staff members to reflect on developments since 2015. Their experiences are relevant to institutions worldwide as they grapple with #BlackLivesMatter and #ShutDownSTEM protests over racial inequalities in society and the colonial foundations of many universities.

Our interviewees reference four key events since the statue fell: the 2018 suicide of Bongani Mayosi, UCT's dean of health sciences; a 2019 report into UCT's institutional culture (see [go.nature.com/3f85j3g](https://go.nature.com/3f85j3g)); the ongoing COVID-19 pandemic; and a [devastating fire last month](#) that gutted the university's African studies reading room and some of South Africa's most treasured histories.

## Of protest and potential

Here is a brief timeline of important events at the University of Cape Town (UCT) in South Africa since the Rhodes Must Fall movement started in 2015.

### March–May 2015

Activist and student Chumani Maxwele hurls a bucket of human faeces over the bronze statue of the nineteenth-century British colonialist Cecil John Rhodes at UCT. This sets off a series of protests, which culminate in the statue's removal on 9 April. The protests then expand to grievances about the curriculum, tuition fees and housing for financially disadvantaged students. On 26 May, vice-chancellor Max Price strikes an agreement with protesting students who had occupied a building on campus, whereby the

students promise to leave the building and UCT will not pursue disciplinary action against the ringleaders.

## October 2015

Protests resurge across South Africa's university campuses. A hashtag made popular earlier in the year, #Feesmustfall, trends once more on Twitter. At UCT, police use stun grenades to disperse students.



South African students protest outside the parliament precinct on 21 October 2015 in Cape Town.Credit: Nardus Engelbrecht/Gallo Images/Getty

## February–March 2016

Protests return to UCT, this time targeting a lack of housing for Black students.

## **6 November 2016**

UCT strikes an agreement with protesting students, granting clemency for protesters. It also establishes a commission to look into the protests of February 2016 and make recommendations on how to transform the institutional culture and address racial and gender inequities.

## **20 March 2019**

The commission publishes its recommendations. The university releases a statement, saying: “We have, reluctantly, concluded that racism does exist at UCT, that it goes beyond attitudes and beliefs and is aided and abetted by poor management systems which administratively result in discrimination on a racial basis.”

## **23 June 2020**

UCT publishes its first ‘transformation benchmarking report’, which will help the university to track its progress on its transformation goals. The report will be published annually.

An inquest into Mayosi’s death found that protests about an increase in university fees in 2016 had placed immense pressure on him as a Black academic leader; students first looked to him as an ally, then tore into him when he supported the university.

And the 2019 report found that racism exists at UCT, and that university management systems contribute to the problem. Critics called the report one-sided, whereas others said that it gave voice to long-standing feelings of not belonging that many Black South Africans experience at the university.

A UCT spokesperson told *Nature* that the mixed reception to the report was “unsurprising”, given the contested nature of the events the report investigated. “After all, UCT is a microcosm of a country where, after many years of confronting this scourge, [racism] remains a serious societal challenge. UCT remains resolute in our determination to defeat it, working

together as a community,” the spokesperson said, adding that many of the report’s recommendations were consistent with initiatives already under way at the university.



Hlumani Ndlovu calls for more role models for Black early-career scientists.

## **HLUMANI NDLOVU: ‘The culture of whiteness remains intransigent’**

When Rhodes fell, I was a postdoctoral fellow in the UCT faculty of health science. I learnt a lot in the weeks leading up to the statue coming down. I was impressed with the solidarity that the students showed. I was also impressed with the intersectionality of their struggle — that it involved not only race, but also gender-based violence and other things that we’ve not tackled as a country.

The death of Mayosi, and the university’s investigation into it, emphasized the burden that is placed on Black staff members. They have to excel academically while also acting as a conduit between the institution and its Black students and staff, and then on top of that, drive the transformation

agenda. That's a huge burden placed on someone who from the start might feel excluded from the university's power systems.

The fall of the Rhodes statue was symbolic, not just for UCT but for our nation and the rest of the world. It symbolized that UCT's structure, with whiteness as the norm, needed to fall — a structure that had become entrenched since UCT was set up as a university for white people. It was a big psychological victory for Black students and for other marginalized groups on campus.



### Career resources for African scientists

But the question now is, has the system fallen? My view is that there's been a desire to transform, but that there have also been people fighting such transformation. The culture of whiteness remains intransigent.

This slow pace of change affects us, as young Black academics. If we don't see Black professors, we don't have anyone to look up to and say they are role models. And role models are important: they can assist you and open doors for you. That's not unique for Black young scientists, but I do think that the need is more acute for us as Black early-career researchers. You can't become what you don't see.

I don't think we as a university community have properly dealt with Mayosi's death. Someone who was a brilliant academic and a big advocate of transformation, a role model, took his own life while leading the most complex faculty at the university. That is bound to have a psychological impact on our desire as Black academics to become leaders.

*Hlumani Ndlovu is an immunologist and senior lecturer in the department of integrative biomedical sciences at UCT.*



PhD student Paballo Chauke is considering leaving academia after he receives his degree because of systemic barriers he has faced. Credit: Paballo Chauke

## **PABALLO CHAUKE: 'My Black body still needs to be accepted'**

I was in the crowd when the Rhodes statue was taken down. It was a watershed moment. My main memory is a feeling of surprise; I never thought racism would be confronted at UCT. When I came to the university in 2010, the feeling among Black South Africans was that you either assimilate into the culture of whiteness or you perish.

I had recently graduated from my honours degree in environmental geography, making ends meet by teaching and working as a research assistant. I come from a poor background; my mother was a domestic worker. When I arrived at UCT, it was the first time that I had come face to face with whiteness. Nearly all my lecturers were white. That said to me that I didn't belong.

After Rhodes fell, I went to the University of Oxford, UK, for my master's in biodiversity, conservation and management. I was one of only four Black people in the 600 students in my college. When I came back to UCT in 2017 for my PhD programme, some things had changed, but much remained the same. 'Transformation' groups had been formed to redress the racial and gender balance of the university, but many of them were led by white women. That bothered me.

Looking back, I realize that I experienced a lot of racism as an undergraduate. Much of it was hidden, not overt. Since Rhodes fell, UCT has hired more Black people, and the current vice-chancellor is Black. But it's like a cappuccino: on top, you have white foam, with a few chocolate flakes sprinkled on top; the bottom is where the dark coffee is. So much needs to change still.

After my PhD, I don't think I want to be an academic. There are so many barriers to jump. My Black body still needs to be accepted by white academics. It's not just about whiteness: I'm queer, I'm poor, I support several of my family members — I need to earn a salary. There's more job security in other careers. Of course, there's the feeling of 'If I don't change it, who will?', but self-care matters, and I'm exhausted.

Still, I don't think UCT should be thought of as a place that can't belong to Black people. When the fire hit, some people on Twitter celebrated that

colonial symbols had burnt. That made me so heartsore and angry. UCT might have its challenges, but it's also a home for Black scholarship.

*Paballo Chauke is a PhD student in environmental geographical science at UCT and a training coordinator for H3ABioNet, a Pan African Bioinformatics Network for the Human Heredity and Health in Africa (H3Africa) consortium.*



UCT lecturer Shannon Morreira says that her unit is placing more emphasis on the skills that students bring, such as language.Credit: Shannon Morreira

## **SHANNON MORREIRA: ‘Changes are happening, but slowly’**

When Rhodes fell, I was already a lecturer in the UCT humanities education-development unit. The unit runs an extended degree programme for people from historically disadvantaged groups in South Africa whose grades would usually be too low to allow them entry. They receive extra academic support and take four years to complete their undergraduate degrees instead of the usual three.

Most students in the programme were Black, and I knew that some were unhappy. So a lot of the conversations that surfaced at UCT about representation and redress after the Rhodes protest were the same ones that we had been having with students earlier. I wasn't surprised when the protests broke out. What did surprise me was the extent to which the grievances were taken up across the university.

As a white staff member teaching predominantly Black students, more than ever before in my life, I could not escape my whiteness. That experience was valuable and valid, but extremely challenging.

Since then, my unit has changed the way it does things. There is now a sense that the unit should focus more on making sure that what is taught, and how, is tailored to students' experiences, rather than expecting them to 'catch up' with UCT's standard learning model.

We are also placing more emphasis on the skills that students bring. South Africa has 17 official languages, and many students are coming in with these huge language resources. We use those resources and allow students to write in languages other than English. We are also shifting content to include more perspectives from the global south in the required reading.

So things are happening. When I began teaching at UCT, research was viewed as much more important than education. Now, that has partially shifted, and teaching and learning are viewed as important, too. Course designers are putting more thought into their target audience than they did before Rhodes fell. As a staff member, I can easily see those changes. But they happen very slowly in the three- or four-year university career of a student, and that can be frustrating for them.

*Shannon Morreira is an anthropologist and senior lecturer in UCT's humanities education-development unit.*



Merlyn Nkomo says that the pandemic has helped her to find her voice.Credit: Linda Nordling

## **MERLYN NKOMO: ‘I feel more settled because I have said my piece’**

When Rhodes fell, I was in the second year of my undergraduate degree in Zimbabwe. Growing up, I had always loved the outdoors and biology. But there are no postgraduate programmes to study conservation in my country, and also none in ornithology, which is my passion. That is why I applied for the conservation master’s programme at UCT’s FitzPatrick Institute of African Ornithology.

I chose UCT because I want to work in conservation projects that interface with African communities. I had hoped that the university would offer a space that recognized African languages and cultures. But when I got here, I

felt I might as well have gone to the United Kingdom to study. Many of the conservation solutions we discuss in the course — such as becoming a vegetarian or vegan to mitigate climate change — are not practical in many African contexts.

My parents don't have a lot of money. My Mandela Rhodes fellowship was my saving grace, because it covers all my costs. I arrived in Cape Town a little over a year ago, just before the pandemic hit. I was one of only two Black people in my class. Then the pandemic lockdown came like a punch in the face. I was already in this space where I didn't feel I belonged, and suddenly I was even more isolated.

Still, the pandemic has given me time to think and write, which I might not have had if I had been preoccupied with the interpersonal aspect of the course. I wrote an article in which I argue that the dearth of Black people studying and practising conservation in Africa is the field's Achilles heel, and that this has to change ([go.nature.com/3eerhae](https://go.nature.com/3eerhae)).

I've engaged with groups that are working to make UCT more inclusive, not just from a racial perspective, but also for people with disabilities. We also want to make the research done at the institute more visible. It's about transforming the space. We're in this old colonial-style building with pictures of old white natural-history collectors on the walls.

I feel more comfortable at UCT than when I arrived, but I don't think it's because the space has changed, or that I have become more adept at navigating it. Rather, I think I feel more settled because I have raised my voice and said my piece. People cannot ignore how I feel as a girl from a colonized country among them.

Six years on from Rhodes falling, I do think people are more mindful and thoughtful about inclusivity. I find that some white people feel like they have to defend the status quo. But nobody alive today is at fault for the system that's in place or guilty of creating it, so nobody should need to defend it. We are on the same side; the world will work better if it becomes less polarized.

*Merlyn Nkomo is a Mandela Rhodes Scholar from Zimbabwe and PhD student in conservation biology at UCT.*

Nature **593**, 465-467 (2021)

doi: <https://doi.org/10.1038/d41586-021-01321-3>

These interviews have been edited for length and clarity.

## **Jobs from Nature Careers**

- - - [All jobs](#)
    - 
    - [Post-Doc - Numerical and microfabrication development of 4D metamaterials for mechanical, acoustic and photonics applications](#)  
[Université Bourgogne Franche-Comté \(UBFC\)](#)  
[BESANCON, France](#)  
[JOB POST](#)
    - [Postdoctoral Position - Ecological Modeler](#)  
[The University of British Columbia \(UBC\)](#)  
[Vancouver, Canada](#)  
[JOB POST](#)
    - [Postdoctoral Fellow | Zandstra Stem Cell Bioengineering Lab](#)

[The University of British Columbia \(UBC\)](#)

[Vancouver, Canada](#)

[JOB POST](#)

- [Personalreferent\\*in \(d/m/w\), Gruppe Personaladministration](#)

[Helmholtz Centre for Heavy Ion Research GmbH \(GSI\)](#)

[Darmstadt, Germany](#)

[JOB POST](#)

[Download PDF](#)

[Download PDF](#)

## Related Articles



[Paths to success for African scientists](#)



[How African scientists can give back to their home continent](#)



[Racism in academia, and why the ‘little things’ matter](#)



- [Meeting the challenges of research across Africa](#)



- [How apartheid's legacy can still cast a shadow over doctoral education in South Africa](#)



- [Collection: Diversity and scientific careers](#)

---

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-021-01321-3>

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

## WHERE I WORK

17 May 2021

# Designing nano-sized chemotherapy

Nanoparticles carrying chemotherapeutic drugs could help people with cancer escape some of the drugs' side effects, hopes Silvia Giordani.

- [Virginia Gewin](#) <sup>0</sup>

1. Virginia Gewin

1. Virginia Gewin is a freelance writer in Portland, Oregon.

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

 Silvia Giordani in her lab at Dublin City University, Ireland.

Silvia Giordani is professor and chair of the nanomaterials department and head of the school of chemical science at Dublin City University in Ireland.  
Credit: Chris Maddaloni for *Nature*

As a teenager, I realized I was too sensitive to suffering to become a medical doctor, yet I still wanted to cure the world. I focused on chemistry for my PhD at the University of Miami, Florida, and then pursued physics as a postdoc at Trinity College Dublin, working on carbon nanotubes. The large surface area of nanomaterials gives them unique optical and electrical properties. I learnt that I could add nanoscale ‘switches’ to direct the nanotubes to perform specific functions in human cells. As toxicity concerns grew over nanotubes because of their rigid, hollow shape, I began

working on carbon nanoparticles, which have those same desirable properties.

I call these devices nano-onions, because they consist of concentric layers of carbon just 5 nanometres in diameter. I knew that they would be perfect for biomedical applications, because they are small enough to enter a human cell and chemically reactive enough to use in imaging, diagnostics or targeted drug delivery. The one problem was making them soluble in water so that the body could absorb them. My colleague and I did that by adding phospholipids and hyaluronic acid. Cancer cells prefer to bind to hyaluronic acid, so that tricks them into absorbing the nano-onions.

In this photo taken in my laboratory, you can see two key pieces of equipment. In the foreground is the furnace, where we synthesize the nano-onions in a process that I patented. Behind me is the fume cupboard, where we can work safely with the toxic chemotherapeutic compounds that we add to the nano-onions. Although we haven't commercialized anything yet, I've worked for a decade to confirm that these nano-onions have no toxic effects.

Ideally, my work will improve targeted delivery of chemotherapeutic drugs and avoid adverse effects on healthy cells. I hope to create nanoparticles that can deliver diagnostic, imaging and therapeutic functions — all together.

Nature **593**, 470 (2021)

doi: <https://doi.org/10.1038/d41586-021-01322-2>

## Jobs from Nature Careers

- - - [All jobs](#)
    - 
    -

**Post-Doc - Numerical and microfabrication development of 4D metamaterials for mechanical, acoustic and photonics applications**

Université Bourgogne Franche-Comté (UBFC)

BESANCON, France

JOB POST

**▪ Postdoctoral Position - Ecological Modeler**

The University of British Columbia (UBC)

Vancouver, Canada

JOB POST

**▪ Postdoctoral Fellow | Zandstra Stem Cell Bioengineering Lab**

The University of British Columbia (UBC)

Vancouver, Canada

JOB POST

**▪ Personalreferent\*in (d/m/w), Gruppe Personaladministration**

Helmholtz Centre for Heavy Ion Research GmbH (GSI)

Darmstadt, Germany

JOB POST

Download PDF

[Download PDF](#)

## Related Articles



- [Three researchers who are coming at cancer from all angles](#)



- [On the trail of travelling RNA](#)



- [Taking graphene out of the laboratory and into the real world](#)

---

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-021-01322-2>

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

# Research

- **Previously unknown type of protein crosslink discovered**

[ 05 May 2021]

News & Views • Molecular crosslinks known as disulfides stabilize the 3D structures of many proteins, and sometimes regulate protein function. But disulfides are not alone — another type of regulatory protein crosslink has been discovered.

- **Ultracold chemical reactions reveal the quantum mechanism of product formation** [ 19 May 2021]

News & Views • Quantum chemistry is challenging to model computationally. An ultracold chemical reaction has now been used to test models with great precision, providing a benchmark for future quantum-chemistry calculations.

- **Revolutionary view of two ways to split a mitochondrion**

[ 05 May 2021]

News & Views • Organelles called mitochondria divide in at least two contexts: during cell growth and in response to mitochondrial damage. The finding that division is different in these two contexts sheds light on the regulatory pathways involved.

- **Conversations, and how we end them** [ 28 April 2021]

News & Views • How we feel about the duration of our conversations has rarely been studied. New research has asked people about the lengths of their conversations, and whether they end when they want them to.

- **Iron and nickel vapours are present in most comets** [ 19 May 2021]

News & Views • The detection of iron and nickel vapours in a broad range of Solar System comets, and of nickel vapour in a comet from outside the Solar System, provides a glimpse into the organic chemistry of young planetary systems.

- **Unconventional viral gene expression mechanisms as therapeutic targets** [ 19 May 2021]

Review Article • This Review outlines the gene and protein expression strategies used by viruses to expand the efficiency of their coding and regulatory sequences, and the implications of these mechanisms for developing antiviral agents.

- **The data-driven future of high-energy-density physics** [ 19 May 2021]

Perspective • This Perspective discusses how high-energy-density physics could tap the potential of AI-inspired algorithms for extracting relevant information and how data-driven automatic control routines may be used for optimizing high-repetition-rate experiments.

- **Iron and nickel atoms in cometary atmospheres even far from the Sun** [ 19 May 2021]

Article • High-resolution ultraviolet and optical spectra of a large sample of comets show that Fe i and Ni i lines are ubiquitous, even when the comets are far from the Sun.

- **Gaseous atomic nickel in the coma of interstellar comet 2I/Borisov** [ 19 May 2021]

Article • Atomic nickel vapour is found in the cold coma of the interstellar comet 2I/Borisov at a distance of 2.3 astronomical units, equivalent to an equilibrium temperature of 180 kelvin.

- **Precision test of statistical dynamics with state-to-state ultracold chemistry** [ 19 May 2021]

Article • The chemical reaction  $2\text{KRb} \rightarrow \text{K}_2 + \text{Rb}_2$  is studied under ultralow temperatures at the quantum state-to-state level, allowing unprecedented details of the reaction dynamics to be observed.

- **Tunable non-integer high-harmonic generation in a topological insulator** [ 19 May 2021]

Article • High-harmonic generation from the Dirac-like surface state of a topological insulator is separated from bulk contributions and continuously tuned by the carrier-envelope phase of the driving lightwave.

- **Discovery, characterization and engineering of ligases for amide synthesis** [ 19 May 2021]

Article • A family of enzymes—coronafacic acid ligases, involved in the synthesis of bacterial phytotoxins—are found to catalyse amide bond formation with a wide range of substrates.

- **Overwintering fires in boreal forests** [ 19 May 2021]

Article • Large forest fires in Alaska and the Northwest Territories can ‘overwinter’ and then reignite in the following fire season, contributing up to one-third of the burned area in individual years.

- **Somatic mutation landscapes at single-molecule resolution**  
[ 28 April 2021]

Article • NanoSeq is used to detect mutations in single DNA molecules and analyses show that mutational processes that are independent of cell division are important contributors to somatic mutagenesis.

- **Mouse prefrontal cortex represents learned rules for categorization** [ 21 April 2021]

Article • Neurons in the mouse medial prefrontal cortex acquire category-selective responses with learning.

- **Clofazimine broadly inhibits coronaviruses including SARS-CoV-2** [ 16 March 2021]  
Article • The anti-leprosy drug clofazimine inhibits coronavirus replication in several cell models and shows potent antiviral activity against SARS-CoV-2 infection in a hamster model, particularly when used in combination with remdesivir.
- **Bispecific IgG neutralizes SARS-CoV-2 variants and prevents escape in mice** [ 25 March 2021]  
Article • The bispecific IgG1-like CoV-X2 prevents SARS-CoV-2 spike binding to ACE2, neutralizes SARS-CoV-2 and its variants of concern, protects against disease in a mouse model, whereas the parental monoclonal antibodies generate viral escape.
- **In vivo CRISPR base editing of PCSK9 durably lowers cholesterol in primates** [ 19 May 2021]  
Article • In a cynomolgus macaque model, CRISPR base editors delivered in lipid nanoparticles are shown to efficiently and stably knock down PCSK9 in the liver to reduce levels of PCSK9 and low-density lipoprotein cholesterol in the blood.
- **Distinct fission signatures predict mitochondrial degradation or biogenesis** [ 05 May 2021]  
Article • Mitochondrial fission at the organelle periphery generates small daughter mitochondria that are removed by mitophagy whereas fission at the midzone leads to proliferation.
- **Neuronal enhancers are hotspots for DNA single-strand break repair** [ 25 March 2021]  
Article • DNA single-strand breaks in neurons accumulate at high levels in functional enhancers.
- **Structure and dynamics of a mycobacterial type VII secretion system** [ 12 May 2021]  
Article • A cryo-electron microscopy structure of the inner membrane complex of the ESX-5 type VII secretion system of *Mycobacterium tuberculosis* reveals an important role of interactions with MycP5 protease for complex integrity.
- **Structure of human telomerase holoenzyme with bound telomeric DNA** [ 21 April 2021]  
Article • A high-resolution structure of human telomerase bound to telomeric DNA reveals details of telomerase assembly and its active site, and sheds light on how mutations alter telomerase function.
- **Structures of telomerase at several steps of telomere repeat synthesis** [ 12 May 2021]

Article • Cryo-electron microscopy structures of Tetrahymena telomerase with telomeric DNA at several steps of nucleotide addition provide insights into the structural basis of telomere repeat synthesis.

- **A lysine–cysteine redox switch with an NOS bridge regulates enzyme function** [ 05 May 2021]

Article • A NOS bridge between cysteine and lysine residues serves as an allosteric redox switch in the transaldolase enzyme of *Neisseria gonorrhoeae*, demonstrating the existence of protein regulatory switches that contain covalent crosslinks other than disulfides.

## NEWS AND VIEWS

05 May 2021

# Previously unknown type of protein crosslink discovered

Molecular crosslinks known as disulfides stabilize the 3D structures of many proteins, and sometimes regulate protein function. But disulfides are not alone — another type of regulatory protein crosslink has been discovered.

- [Deborah Fass](#) <sup>0</sup> &
- [Sergey N. Semenov](#) <sup>1</sup>

### 1. [Deborah Fass](#)

1. Deborah Fass is in the Department of Chemical and Structural Biology, Weizmann Institute of Science, Rehovot 7610001, Israel.

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

### 2. [Sergey N. Semenov](#)

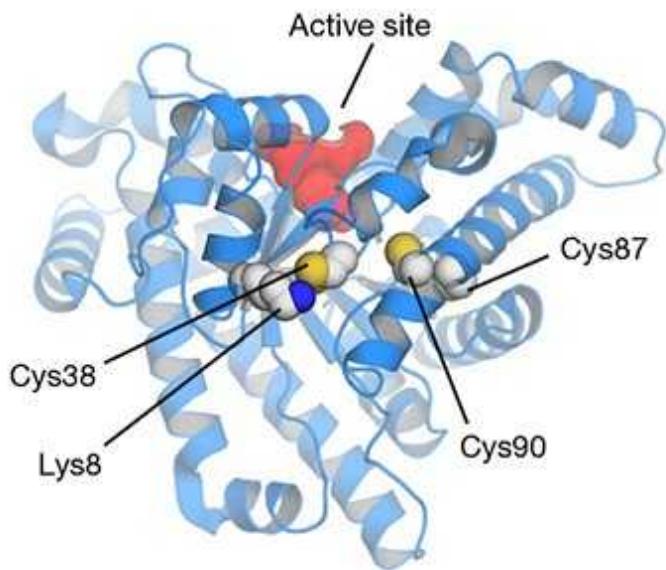
1. Sergey N. Semenov is in the Department of Molecular Chemistry and Materials Science, Weizmann Institute of Science.

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

The characteristic shapes and functions of proteins depend on the order in which the constituent amino acids are linked into chains. However, further chemical modifications are often made after the amino acids are strung together. These modifications include crosslinks between certain amino-acid

residues. The most common type of crosslink is a disulfide: two sulfur atoms connected by a covalent bond. [Writing in Nature](#), Wensien *et al.*<sup>1</sup> report an entirely different type of protein crosslink, in which an oxygen atom connects a nitrogen to a sulfur atom. Moreover, the authors present evidence that such N–O–S bridges have gone unnoticed in previously reported structural analyses of other proteins.



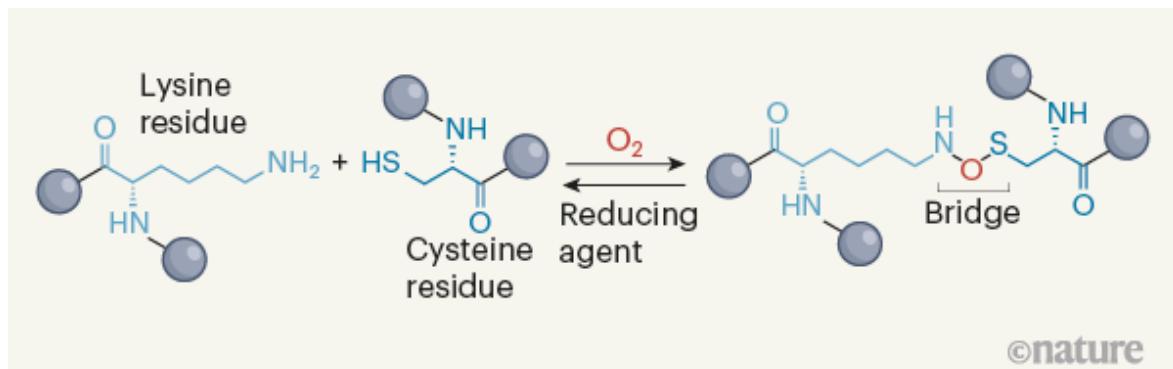
[Read the paper: A lysine–cysteine redox switch with an NOS bridge regulates enzyme function](#)

Wensien and colleagues were studying the enzyme transaldolase from *Neisseria gonorrhoeae*, a bacterium that causes the eponymous sexually transmitted disease. They observed that the purified enzyme was almost inactive, but that activity could be restored by using reducing agents commonly used to break disulfide bonds.

To form a disulfide, the side chains of two cysteine amino-acid residues — which can be far apart along the protein chain — must come close together in space. Noticing that the transaldolase contains a few cysteine residues, the authors reasoned that two of them might form a disulfide that inactivates the enzyme. However, when they replaced each of the cysteines individually with another amino acid, they found that only one of the resulting mutant enzymes resisted inactivation. If the enzyme really did contain an inactivating disulfide, then replacement of either of the two participating

cysteines would be expected to have this effect (although there are exceptions to this simple generalization<sup>2</sup>). So what was going on?

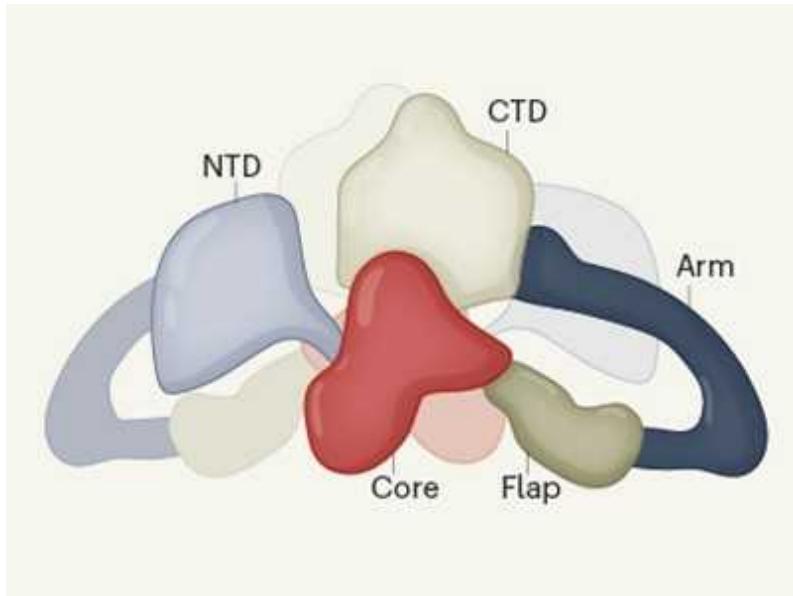
Wensien *et al.* solved the puzzle by using X-ray crystallography to determine the structure of the transaldolase at atomic resolution. This analysis revealed a covalent connection between a cysteine and a lysine residue — an N–O–S bridge — instead of a cysteine–cysteine disulfide crosslink (Fig. 1). Intriguingly, the oxygen atom in the N–O–S bridge does not arise from the groups in the side chains of either of those residues (the cysteine side chain contains a thiol (SH) group, whereas the lysine side chain contains an amine ( $\text{NH}_2$ ) group). However, the authors saw an oxygen molecule close to these side chains in the crystal structure of the reduced form of the protein (which lacks the crosslink). This observation supports the authors' reasonable speculation that an oxygen molecule contributes an oxygen atom to the N–O–S bridge.



**Figure 1 | Formation of a bridge between lysine and cysteine amino-acid residues.** Wensien *et al.*<sup>1</sup> report that the transaldolase enzyme of the *Neisseria gonorrhoeae* bacterium contains a molecular crosslink in which an oxygen atom bridges a nitrogen in the side chain of a lysine amino-acid residue and a sulfur atom in the side chain of a cysteine residue. The authors propose that such N–O–S bridges are formed by the reaction of the side chains with an oxygen molecule ( $\text{O}_2$ ). The N–O–S motif is ordinarily unstable, and it is presumably stabilized by the surrounding protein structure. However, it can be cleaved by reducing agents. Spheres represent other segments of the protein molecule.

The discovery of an N–O–S bridge in proteins is noteworthy because non-biological reactions that produce such a molecular motif are not known, with the possible exception of one unusual molecule (a type of cyclic aromatic compound<sup>3</sup>). The formation of an N–O bond in small-molecule chemistry requires strong oxidizing conditions<sup>4</sup>, but such conditions would probably also convert sulfur atoms to higher oxidation states than that of the sulfur in the N–O–S bridge. Furthermore, small molecules containing the N–O–S motif might be in danger of undergoing disproportionation — a process in which two of the same molecules react with one another to produce two different products. In the context of a protein, favourable positioning of the cysteine’s thiol, the lysine’s amine and an oxygen molecule might aid the oxidation needed for N–O–S formation, whereas spatial (steric) constraints imposed by the surrounding protein structure might stabilize the crosslink and block further oxidation of the sulfur.

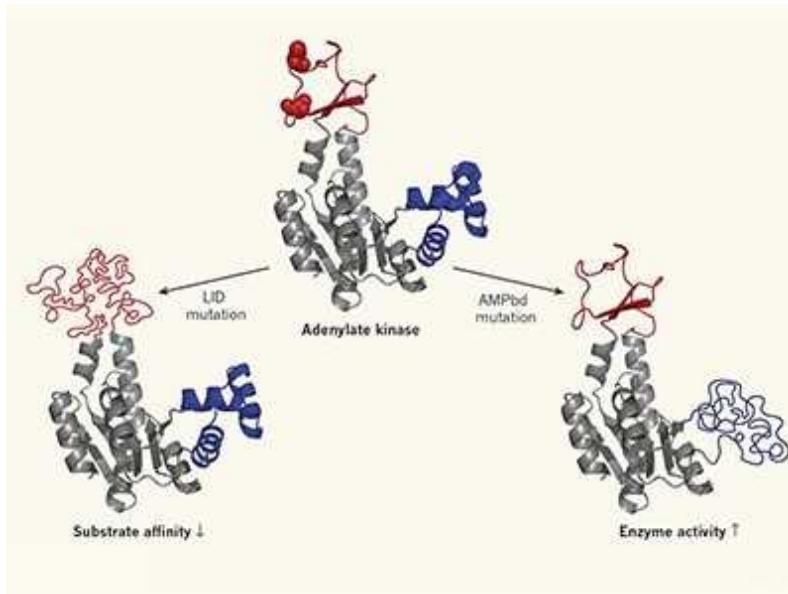
The authors speculate about possible mechanisms for bridge formation, favouring a reaction in which hydroxyl groups (OH) are added to both the sulfur atom of the cysteine and the amine group of the lysine side chain (see Extended Data Fig. 3b of ref. 1). Given the novelty of the N–O–S crosslink and the focus of the study on structural biology, important details of the chemical mechanism remain to be addressed. Specifically, the way in which the oxygen molecule is activated to take part in this reaction is not described. Many chemical reactions with molecular oxygen involve free radicals<sup>5</sup>, and so radical pathways should be explicitly considered in mechanisms for generating the N–O–S bridge.



### Molecular architecture of the key precursor of thyroid hormones revealed

Another mechanistic issue is how the crosslink affects enzymatic activity. The structure of the catalytic site of the crosslinked transaldolase differs only slightly from that of the non-crosslinked version of the enzyme. How formation of the N–O–S bridge inhibits catalysis is thus not obvious. The authors focused their attention on these minor structural differences, but they also observed that the crosslinked transaldolase is more resistant to heat-induced unfolding than is the non-crosslinked version. This result is not surprising and implies that the crosslinked enzyme undergoes fewer conformational fluctuations that could lead to unfolding. Smaller-scale fluctuations could be required for catalytic activity, and might also be restrained by the presence of the crosslink.

There is a key conceptual difference between the use of disulfide bonds for regulating protein function and the use of an N–O–S bridge. The formation of disulfide bonds is chemically reversible, which means that, in biological systems, disulfides are frequently made and broken in ‘exchange’ reactions with other molecules that contain disulfides or thiols. By contrast, the N–O–S linkage is formed by a different chemistry from that by which it is cleaved — that is, molecular oxygen is used to form the bridge but is not released when the bridge is broken.



### Enzymes can adapt to cold by wiggling regions far from their active site

Moreover, the thermodynamics of these reactions indicate that it is difficult to form the N–O–S bridge but easy to break it. N–O–S crosslinks therefore might have evolved to enable the selective activation of an enzyme under conditions in which disulfides are preserved. The particular advantages conferred by the N–O–S bridge, and the biological scenarios in which it is more useful than a disulfide, can now be explored.

The discovery of a new protein linkage has implications beyond the specifics of the enzyme studied and of the N–O–S crosslink itself. Perhaps surprisingly, the task of generating structural models for proteins is sometimes more difficult when high-resolution X-ray data are available. Variations in protein conformation or in chemical composition might be buried in the noise of the data at low resolution, but this heterogeneity becomes visible at high resolution and must therefore be interpreted<sup>6</sup>. Unanticipated chemical groups might also be lurking in the data. Wensien and colleagues' study will inspire structural biologists to investigate deviations from expectation in their electron-density maps of biomolecules.

It has long been appreciated that enzymes are the world's best organic chemists, because they can promote reactions that would be almost impossible in their absence. The findings show that the covalent chemistry of enzymes themselves can also defy chemical intuition.

Nature **593**, 343–344 (2021)

doi: <https://doi.org/10.1038/d41586-021-01135-3>

## References

1. 1.

Wensien, M. *et al. Nature* **593**, 460–464 (2021).

2. 2.

Mor-Cohen, R. *Antioxid. Redox. Signal.* **24**, 16–31 (2016).

3. 3.

Hassan, A., Ibrahim, Y. R. & Shawky, A. M. *J. Sulfur Chem.* **28**, 211–222 (2007).

4. 4.

Challis, B. C. & Butler, A. R. in *The Chemistry of the Amino Group* (ed. Patai, S.) Ch. 6 (Interscience, 1968).

5. 5.

Sawyer, D. T. *Oxygen Chemistry* Ch. 5–6 (Oxford Univ. Press, 1991).

6. 6.

Lang, P. T. *et al. Protein Sci.* **19**, 1420–1431 (2010).

## Jobs from Nature Careers

- - - [All jobs](#)
    -

- **Post-Doc - Numerical and microfabrication development of 4D metamaterials for mechanical, acoustic and photonics applications**

Université Bourgogne Franche-Comté (UBFC)

BESANCON, France

JOB POST

- **Postdoctoral Position - Ecological Modeler**

The University of British Columbia (UBC)

Vancouver, Canada

JOB POST

- **Postdoctoral Fellow | Zandstra Stem Cell Bioengineering Lab**

The University of British Columbia (UBC)

Vancouver, Canada

JOB POST

- **Personalreferent\*in (d/m/w), Gruppe Personaladministration**

Helmholtz Centre for Heavy Ion Research GmbH (GSI)

Darmstadt, Germany

JOB POST

Download PDF

[Download PDF](#)

## Related Articles



- [\*\*Read the paper: A lysine–cysteine redox switch with an NOS bridge regulates enzyme function\*\*](#)



- [\*\*Enzymes can adapt to cold by wiggling regions far from their active site\*\*](#)



- [\*\*Molecular architecture of the key precursor of thyroid hormones revealed\*\*](#)

- [\*\*See all News & Views\*\*](#)

---

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-021-01135-3>

## NEWS AND VIEWS

19 May 2021

# Ultracold chemical reactions reveal the quantum mechanism of product formation

Quantum chemistry is challenging to model computationally. An ultracold chemical reaction has now been used to test models with great precision, providing a benchmark for future quantum-chemistry calculations.

- [Nandini Mukherjee](#) <sup>0</sup>

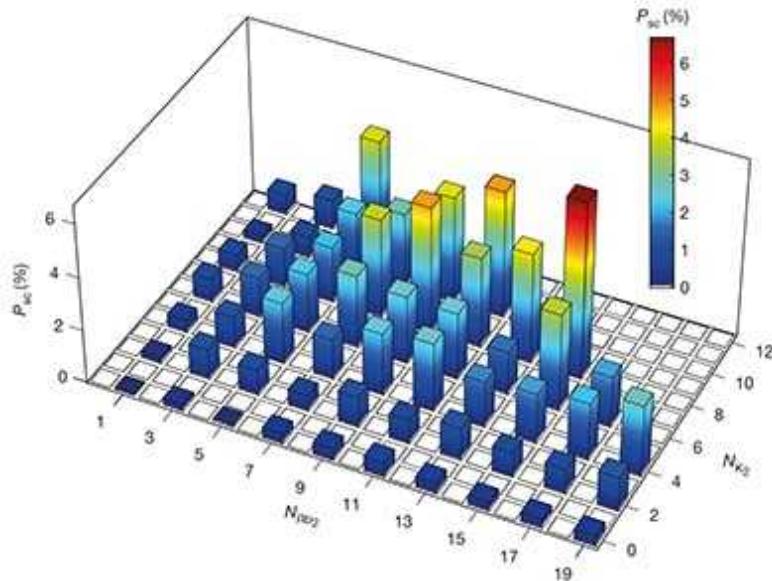
### 1. [Nandini Mukherjee](#)

1. Nandini Mukherjee is in the Department of Chemistry, Stanford University, Stanford, California 94305, USA.

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

Researchers have long dreamt of following chemical reactions from a single quantum state of the reactant molecules to all the possible quantum states of the products, to understand the quantum dynamics that drive chemistry at the most fundamental level. The invaluable data obtained by such an experiment could be compared directly with theoretical results to work out the quantum mechanism of reactions. [Writing in Nature](#), Liu *et al.*<sup>1</sup> report that this dream has come true for a reaction of ultracold diatomic molecules consisting of one potassium atom and one rubidium atom (KRb). By mapping the statistical distribution of all 57 of the possible quantum states of the reaction products, the authors were able to establish the validity of quantum-statistical models of reactions.



[Read the paper: Precision test of statistical dynamics with state-to-state ultracold chemistry](#)

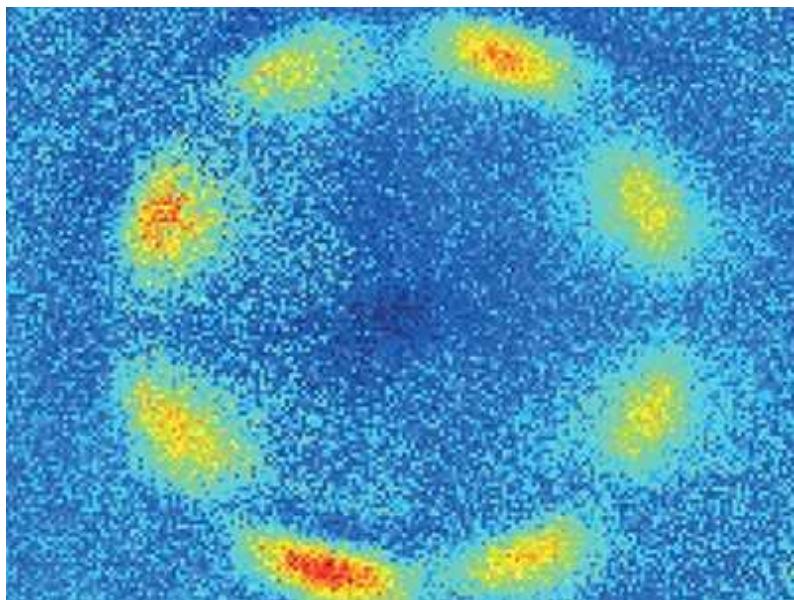
In the past few years, it has become possible to prepare dense samples of ultracold diatomic molecules of the alkali metals (those in group 1 of the periodic table) in their absolute ground states<sup>2,3</sup>. This means that the molecules are in their lowest electronic, vibrational and rotational energy states, and the translational kinetic energy of each molecule — the energy associated with the movement of the molecule's centre of mass through space — is vanishingly small (less than 1 microkelvin, or about  $10^{-10}$  electronvolts).

The ability to prepare such molecules has opened up the possibility of studying chemical reactions in which the reactants start in a single internal quantum state (that is, they have identical electronic, vibrational and rotational states) and collide with a translational energy of less than  $10^{-10}$  eV. Such reactions are an ideal test bed for validating quantum-dynamical models of reactions. Once validated and benchmarked, quantum-dynamical calculations could be used to model more-complex reactions. The reaction in which two KRb molecules exchange atoms to form K<sub>2</sub> and Rb<sub>2</sub> is particularly suitable for this purpose, because it is exoergic — the reaction releases energy. This implies that the reaction can proceed when the

reactants are in their absolute ground state and have a near-zero translational kinetic energy.

The energy released when two molecules collide and react is partitioned between the various internal energy states and the translational motion of the products. The challenge in computational chemistry is to determine the reaction pathways that connect the initial reactant state to a specific product state, through a complex intermediate state. The probabilities with which product molecules in different quantum states are formed from a single reactant state must therefore be determined experimentally.

But, until the past few years, ultracold reactions were characterized only by the rate at which reactant molecules are lost from a trap as a result of scattering from reactive and non-reactive collisions<sup>4</sup>. The states of the product molecules were not characterized, and so the precise reaction pathways involved were obscured. Measuring the distribution of products between all the possible quantum states that can be produced in an ultracold reaction has remained an unresolved challenge.



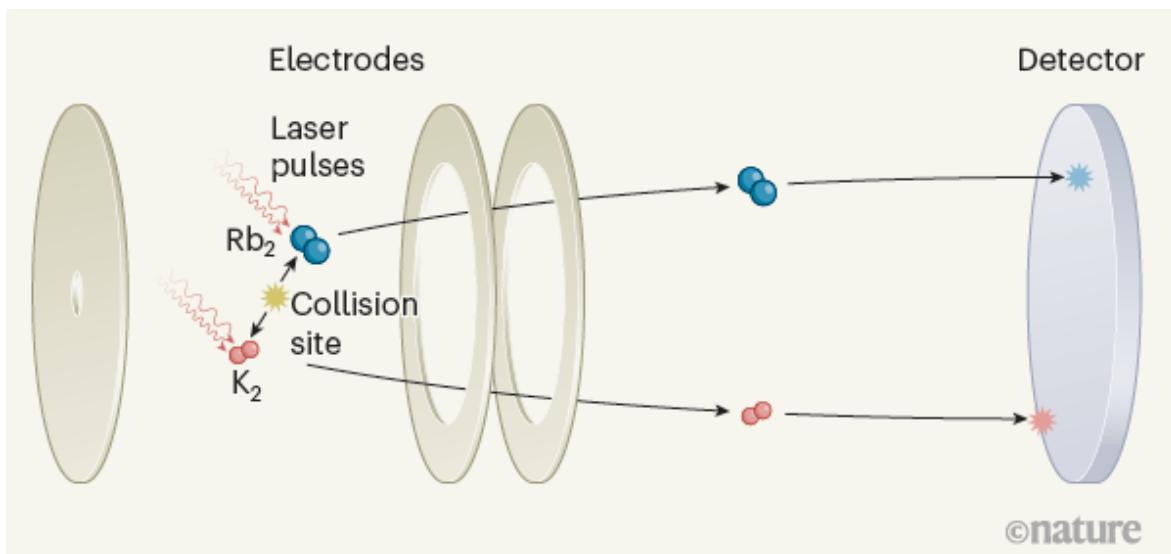
### Quantum control of light-induced reactions

A mass-spectrometry technique that can selectively analyse molecules in particular quantum states and map their velocities has been extensively developed for experiments with beams of molecules<sup>5,6</sup>. A team of

researchers, including some of the authors of the current paper, previously adapted this technique to study KRb molecules in an ultracold reaction vessel<sup>7</sup>. However, the small sample size and large number of possible product states make it much more challenging to measure the distribution of these states than in the molecular-beam experiments, which used lighter molecules that have fewer potential product states.

The energy released when two ultracold KRb molecules exchange atoms is partitioned only between the rotational and translational motion of the product molecules — the energy released in the reaction is insufficient to excite vibrational motion. The joint quantum state of the products is designated by the quantum numbers ( $N_{K_2}$  and  $N_{Rb_2}$ ) that define the rotational energies of  $K_2$  and  $Rb_2$ . In their experiments, Liu *et al.* developed a scheme for detecting the rotational states of pairs of  $K_2$  and  $Rb_2$  molecules that were produced simultaneously in the ultracold reaction. However, to determine the states of the product pairs accurately, the authors also needed to work out whether simultaneously detected  $K_2$  and  $Rb_2$  molecules were actually produced in the same reaction event.

Liu and colleagues solved this problem using the principle of conservation of linear momentum.  $K_2$  and  $Rb_2$  molecules that originate from a single collision of two KRb molecules fly apart with equal and opposite momenta (Fig. 1). The positions of the two product ions imaged in the experiments therefore correlate with each other. By detecting such correlations, the authors identified pairs of molecules produced from the same reaction events, and thereby accurately determined the statistical distribution of products between all 57 possible joint quantum states.

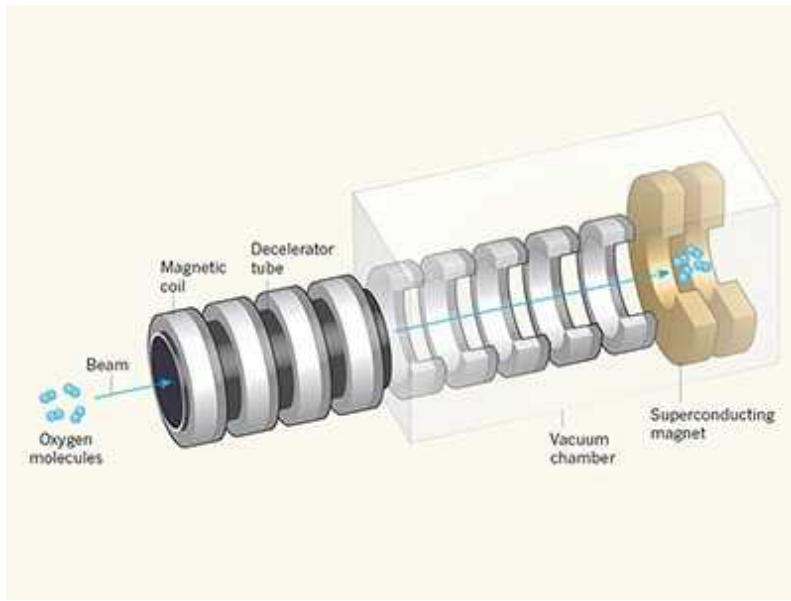


**Figure 1 | Detection of products formed from the collision of ultracold molecules.** Liu *et al.*<sup>1</sup> analysed ultracold reactions in which KRb molecules (K, potassium; Rb, rubidium) are produced in a single quantum state (the absolute ground state), and collide to produce K<sub>2</sub> and Rb<sub>2</sub> molecules. The products of one reaction event are shown rebounding from a collision site on the left. The rotational states of K<sub>2</sub> and Rb<sub>2</sub> molecules were detected by simultaneously and selectively ionizing them with frequency-tuned laser pulses, and the electric fields produced by a series of circular electrodes propelled the ions towards a position-sensitive detector. K<sub>2</sub> and Rb<sub>2</sub> molecules that are produced by the same collision have equal and opposite momenta, which means that the positions at which they strike the detector are correlated. The authors identified such correlated pairs of ions and analysed the quantum state of the product ions. In this way, they tracked the statistical distribution of all 57 of the quantum states that can be produced from a single quantum state of the reactants.

The exact quantum-dynamical calculations needed to describe this type of reaction are currently intractable, but a purely quantum-statistical model can be used to predict the probability that a reaction will lead to a particular quantum state<sup>8</sup>. According to this model, the probability depends on the state's degeneracy — the number of equal-energy versions of a product state that, in this case, is characterized by the rotational energies of K<sub>2</sub> and Rb<sub>2</sub>. Liu and colleagues' experiments do not distinguish between these degenerate states, and so the measured counts represent the probability of

finding  $K_2$  and  $Rb_2$  in all of the degenerate states that have an energy defined by a given combination of  $N_{K_2}$  and  $N_{Rb_2}$ . By comparing the measured counts with the reaction probability calculated from the degeneracy of a state using the statistical model, the authors show that their measurements largely agree with the predictions of the model.

Liu *et al.* observed deviations from the statistical predictions for the limiting ‘threshold’ case, in which the energy released by the reaction is almost completely transferred to the rotational motion of the product molecules, leaving only a small amount of translational kinetic energy for the products to fly apart. In this scenario, the centrifugal energy that arises from the orbiting motion of the product molecules about their common centre of mass gives rise to a centrifugal energy barrier that tends to prevent them from separating into free product states. This reduces the likelihood that degenerate states that have high orbital angular momentum will occur in a product state. That violates the statistical model, which assumes that all degenerate orbital states belonging to a product state have an equal probability of forming.



### [A versatile cold-molecule collider](#)

The authors’ ability to precisely control the quantum states of the reactants in their experiments allowed them to investigate the threshold states of the products that emerge at a minimum translational kinetic energy of

approximately 1 mK (about  $10^{-7}$  eV). At this low energy, the wave character of matter is manifested, allowing the threshold state to ‘tunnel’ across the centrifugal barrier — this tunnelling explains why the authors’ experimental results deviate from the statistical model. By comparing the tunnelling probability calculated from theory with their measurements, Liu *et al.* determined the translational kinetic energy of the threshold product state, and thereby precisely quantified the overall energy released by the KBr atom-exchange reaction.

Liu *et al.* have thus probed a chemical reaction in the finest detail. It will now be interesting to study how the outcome of a reaction is controlled by initial states other than the absolute ground state. For example, reactants could be prepared in a specific rotational–vibrational state using laser excitation<sup>9</sup>; or the orientation of reactants could be controlled through the interaction of their electric or magnetic dipoles with external fields<sup>10</sup>; or reactants could be produced in states of quantum ‘entanglement’<sup>11</sup>. Likewise, external fields could be used to control which product states are accessible. Such highly controlled, state-resolved reactions will serve as a potential benchmark for future theoretical modelling of quantum chemistry.

Will it be possible to exert such fine quantum control over reactions that are of more general interest in chemistry? Substantial progress has been made in developing techniques for cooling a diverse range of molecules to prepare them in stable quantum states<sup>12</sup>. However, so far it has been possible to cool only a few types of molecule to microkelvin temperatures or lower. Nevertheless, Liu and colleagues’ experiment opens up exciting prospects for the study of single-state driven chemical reactions.

Nature **593**, 344–345 (2021)

doi: <https://doi.org/10.1038/d41586-021-01264-9>

## References

1. 1.

Liu, Y. *et al.* Nature **593**, 379–384 (2021).

2. 2.

Moses, S. A., Covey, J. P., Miecnikowski, M. T., Jin, D. S. & Ye, J. *Nature Phys.* **13**, 13–20 (2017).

3. 3.

Bell, M. T. & Softley T. P. *Mol. Phys.* **107**, 99–132 (2009).

4. 4.

Ospelkaus, S. *et al.* *Science* **327**, 853–857 (2010).

5. 5.

Feldman, D. L., Lengel, R. K. & Zare, R. N. *Chem. Phys. Lett.* **52**, 413–417 (1977).

6. 6.

Greaves, S. J., Rose, R. A. & Orr-Ewing, A. J. *Phys. Chem. Chem. Phys.* **12**, 9129–9143 (2010).

7. 7.

Hu, M.-G. *et al.* *Science* **366**, 1111–1115 (2019).

8. 8.

González-Lezana, T. *Int. Rev. Phys. Chem.* **26**, 29–91 (2007).

9. 9.

Bergmann, K. *et al.* *J. Phys. B* **52**, 202001 (2019).

10. 10.

Ni, K.-K. *et al.* *Nature* **464**, 1324–1328 (2010).

11. 11.

Bloch, I. *Nature* **453**, 1016–1022 (2008).

12. 12.

Stuhl, B. K., Hummon, M. T. & Ye, J. *Annu. Rev. Phys. Chem.* **65**, 501–518 (2014).

## Competing Interests

The author declares no competing interests.

## Jobs from Nature Careers

- - - [All jobs](#)
    - 
    - [Post-Doc - Numerical and microfabrication development of 4D metamaterials for mechanical, acoustic and photonics applications](#)  
Université Bourgogne Franche-Comté (UBFC)  
BESANCON, France  
[JOB POST](#)
    - [Postdoctoral Position - Ecological Modeler](#)  
The University of British Columbia (UBC)  
Vancouver, Canada  
[JOB POST](#)
    -

**Postdoctoral Fellow | Zandstra Stem Cell  
Bioengineering Lab**

The University of British Columbia (UBC)

Vancouver, Canada

JOB POST

- **Personalreferent\*in (d/m/w), Gruppe  
Personaladministration**

Helmholtz Centre for Heavy Ion Research GmbH (GSI)

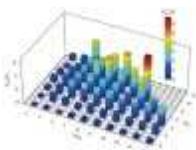
Darmstadt, Germany

JOB POST

Download PDF

Download PDF

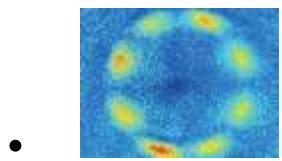
## Related Articles



- **Read the paper: Precision test of statistical dynamics with state-to-state ultracold chemistry**



- **A versatile cold-molecule collider**



- [Quantum control of light-induced reactions](#)
- [See all News & Views](#)

---

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-021-01264-9>

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

NEWS AND VIEWS  
05 May 2021

# Revolutionary view of two ways to split a mitochondrion

Organelles called mitochondria divide in at least two contexts: during cell growth and in response to mitochondrial damage. The finding that division is different in these two contexts sheds light on the regulatory pathways involved.

- [Rajarshi Chakrabarti](#) <sup>0</sup> &
- [Henry N. Higgs](#) <sup>1</sup>

## 1. [Rajarshi Chakrabarti](#)

1. Rajarshi Chakrabarti is at the Geisel School of Medicine at Dartmouth, Hanover, New Hampshire 03755, USA.

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

## 2. [Henry N. Higgs](#)

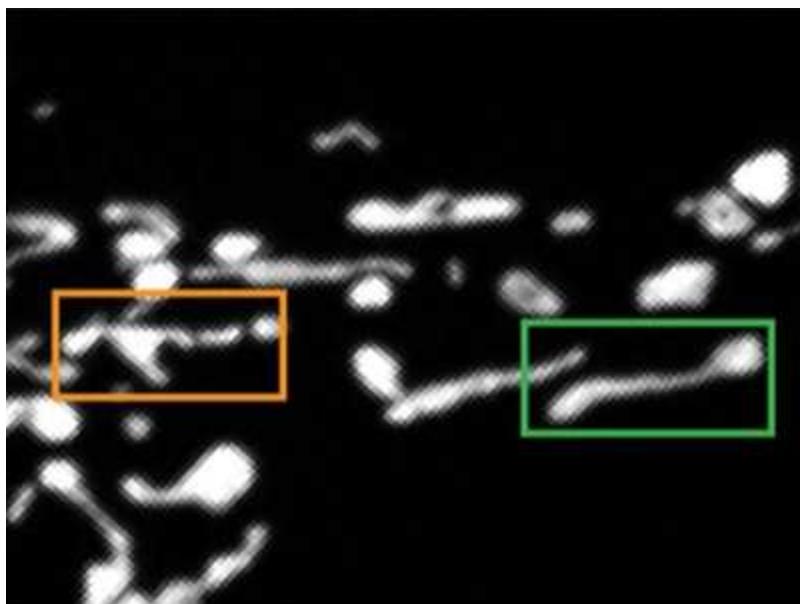
1. Henry N. Higgs is at the Geisel School of Medicine at Dartmouth, Hanover, New Hampshire 03755, USA.

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

Shortly before his life was cut short by the guillotine during the French Revolution, the chemist Antoine Lavoisier made key discoveries about the

biological energy-generating process termed respiration<sup>1</sup>. One of his insights was to realize that respiration is, as he described it<sup>1</sup>, “simply a slow burning of carbon and hydrogen, which is similar to how a lamp or a lighted candle works, and, from that point of view, animals who breathe are veritable flammable bodies who burn and consume themselves”. But how is this ‘burning’ kept under control in cells? [Writing in Nature](#), Kleele *et al.*<sup>2</sup> report some unexpected findings about an organelle at the heart of respiration in animal cells.



[Read the paper: Distinct fission signatures predict mitochondrial degradation or biogenesis](#)

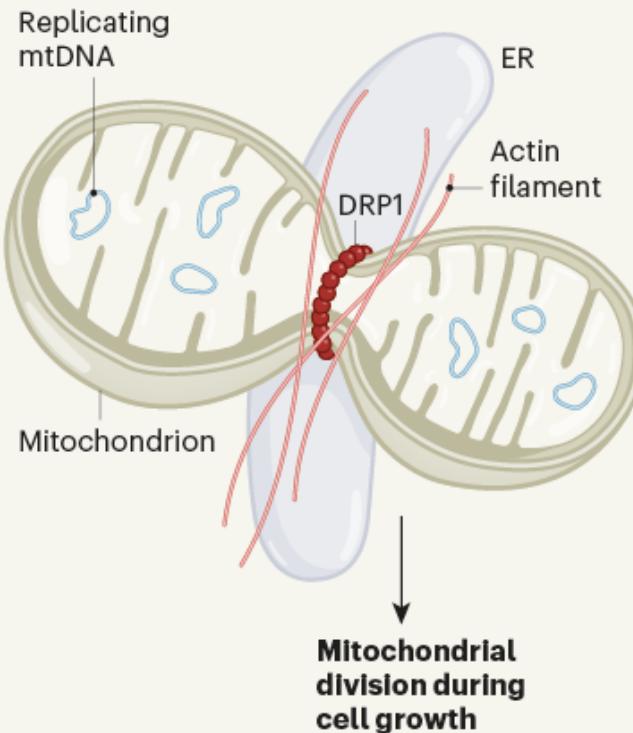
About 150 years after Lavoisier’s time, organelles termed mitochondria were revealed to be where this burning takes place<sup>3,4</sup>, and the mitochondrion is often referred to as the powerhouse of the cell. As with burning, respiration also causes quite a bit of damage, and active mitochondria commonly become defective. Some of the most serious damage that can occur is mutation of the mitochondrial genome, located inside the organelle. A process called mitophagy serves to remove and degrade damaged mitochondria, and is a crucial mechanism for cellular homeostasis. Defects in mitophagy, particularly those affecting long-lived cells such as neurons, are associated with Parkinson’s disease and other neurodegenerative conditions<sup>5</sup>.

During mitophagy, damaged portions of mitochondria separate from healthy portions through mitochondrial division<sup>6</sup>. However, damage is not the only reason for mitochondrial division. It also occurs during cell growth and cell division. In this scenario, the new cellular property generated by cell division is furnished using mitochondria generated by division. In contrast to damage-associated division, mitochondrial division during cell growth is a sign that times are good.

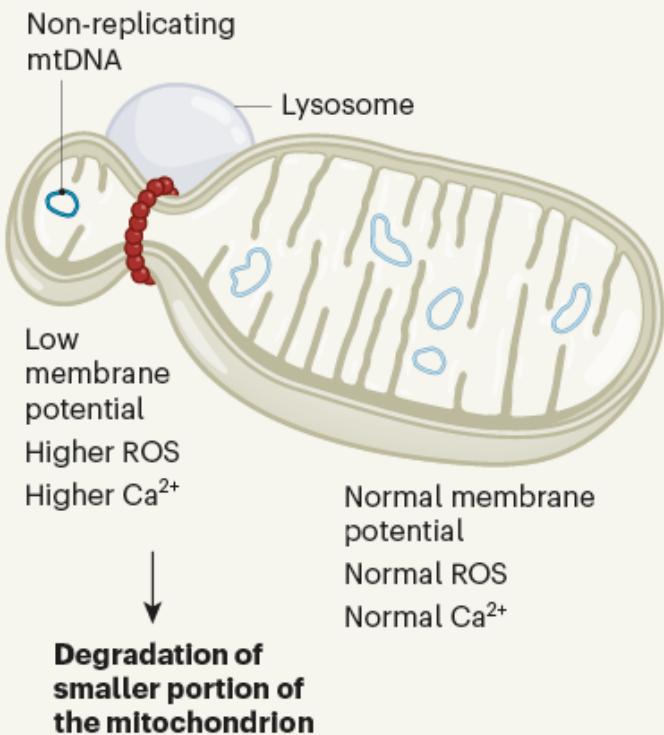
It stands to reason that different mechanisms control mitochondrial division for mitophagy and for cell growth. Although there have been hints of specific types of division, clear evidence has been lacking until now. The protein DRP1 is required for the vast majority of cases of mitochondrial division<sup>6</sup>. DRP1 can be activated in different ways to drive such division in mammals. These include: interaction with mitochondrial DRP1 receptors (MFF, MID49, MID51 and FIS1); DRP1 modification (post-translational alterations); interaction with the actin cytoskeleton (filaments of actin protein) or the mitochondrial lipid cardiolipin; and contact with various organelles, including the endoplasmic reticulum (ER), lysosomes and the Golgi (in the form of Golgi-derived vesicles)<sup>6</sup>. It has been unclear whether these factors contribute to a single division pathway or to different pathways.

Kleele *et al.* conducted careful analysis of mitochondrial division using super-resolution microscopy, and defined two spatially distinct types of division. Midzone division is centrally located on the organelle, whereas peripheral division takes place at the ends of mitochondria (Fig. 1). The two division types occur at similar frequency in Cos-7 cells from monkeys, whereas midzone division is more frequent in mouse neonatal cardiomyocyte cells.

**a Midzone division**



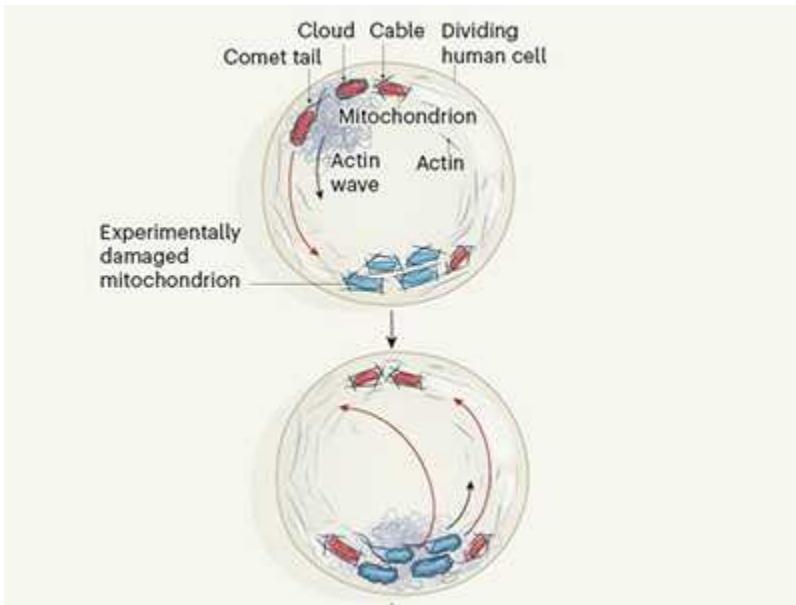
**b Peripheral division**



**Figure 1 | Two pathways for mitochondrial division.** Kleele *et al.*<sup>2</sup> report microscopy studies of organelle division in mammalian cells, which reveal that mitochondria can divide in two ways. **a**, Midzone division is associated with mitochondrial division during cell growth. The organelle divides in the middle, and this process is associated with the protein DRP1, filaments of actin protein, and contact with another organelle — the endoplasmic reticulum (ER). The dividing mitochondrion is healthy and has replicating mitochondrial DNA (mtDNA). **b**, Peripheral division is associated with damaged mitochondria. This division also requires DRP1, but the dividing mitochondrion makes contact with a different organelle, the lysosome. This asymmetric division occurs at the tip of the mitochondrion. The dividing organelle has different properties on either side of the division site in terms of the membrane potential and the level of reactive oxygen species (ROS) and calcium ions ( $\text{Ca}^{2+}$ ). The authors observed that the smaller mitochondrial portion often lacked replicating mtDNA (and in 32% of the divisions it lacked any mtDNA), and that this portion of the organelle was degraded.

The authors demonstrate that peripheral and midzone divisions have substantially different properties. Midzone division occurs in organelles with hallmarks of healthy mitochondria — they do not display signs of abnormalities, such as a reduction of membrane polarization or a change in the level of reactive oxygen species (ROS). By contrast, peripheral division occurs when the tip of the organelle has developed a decrease in membrane potential and an increase in ROS, with a noticeable lack of these alterations in the other portion of the organelle. In addition, this smaller product of a peripheral division often lacks replicating DNA — which is a sign of an unhealthy mitochondrion.

These findings suggest that peripheral division occurs when mitochondria are damaged, and is a precursor to mitophagy. Indeed, the authors report that peripheral divisions increased on exposure to various cellular stresses, and were associated with the accumulation of markers of mitophagy. By contrast, midzone division increased after stimulation of cell proliferation.



### Mitochondria are mixed during cell division

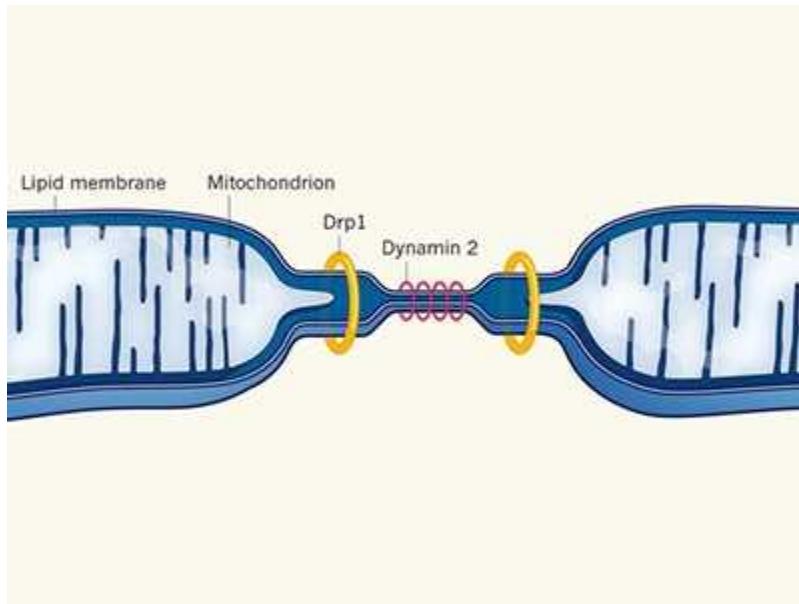
Both types of division are associated with DRP1 accumulation. However, there are differences in other molecular players involved. Midzone division is associated with contact with the ER and with the polymerization of actin filaments through the ER-bound actin-polymerization protein INF2. In addition, the data suggest that MFF has a role in midzone, but not in peripheral, division. Peripheral division is associated with lysosomal contact and with FIS1.

Kleele and colleagues' careful work is valuable, because it clearly demonstrates that there is more than one type of mitochondrial division, thus enabling a more nuanced analysis of division factors based on the reason for division. Moreover, this work is a reminder that we need to walk before we can run when trying to map complicated biological processes such as mitophagy. Otherwise, our understanding of them might be hampered by an incomplete grasp of the earlier processes that lead up to them.

This work also raises exciting questions. Do other factors participate specifically in peripheral or midzone division? In this respect, MID51 and MID49 are particularly interesting because the current work does not provide conclusive results about their role. Other factors worth examining include cardiolipin, Golgi-derived vesicles and post-translational

modifications of DRP1. Another issue to explore is whether cell-type-specific differences make a major contribution, a feature hinted at by the authors' investigation of different cell types.

A fascinating aspect to consider further is the complete compartmentalization of a different profile of calcium, ROS and membrane potential to the smaller portion of a mitochondrion undergoing peripheral division. Different characteristics on either side of the division site have been demonstrated previously for mitochondrial division<sup>7</sup>.



### Double agents for mitochondrial division

One possible mechanism for this compartmentalization is that the inner mitochondrial membrane (the inner of the two membranes surrounding the organelle) undergoes division before the outer membrane, as has been suggested previously<sup>8</sup>. However, compartmentalization in the absence of an independent division of the inner mitochondrial membrane might be possible. This idea is supported by the observation that infoldings of the inner membrane, termed cristae, can maintain membrane potentials that are different from each other, even when in close proximity in a mitochondrion<sup>9</sup>. Another matter to consider is the source of the rising calcium levels in the smaller portion of a peripherally dividing mitochondrion. Calcium transfer from lysosomes is a possibility<sup>10</sup>.

There are some other puzzles. The role of FIS1 in mammalian mitochondrial division has been controversial. Kleele and colleagues' work suggests that FIS1 is the DRP1 receptor for peripheral division, and another study also suggests that FIS1 is a DRP1 receptor<sup>11</sup>. However, other studies<sup>6</sup> indicate that FIS1 depletion has a minimal effect on division, and alternative functions for FIS1 have been described<sup>12,13</sup>. Two explanations for this apparent contradiction are that the other studies on FIS1 were in contexts that did not favour peripheral division, or that the role of FIS1 in peripheral division might be indirect.

Something else to consider is the absence of an increase in mitochondrial calcium levels during midzone division. Previous studies<sup>8,14</sup> have shown that an increase in mitochondrial calcium precedes division events resembling the midzone division described by Kleele and colleagues. It would be interesting to examine the effect of suppressing the mitochondrial calcium uniporter (a protein that pumps calcium across the membrane) on midzone and peripheral division.<sup>[EP]</sup>A final question is whether there are only two types of mitochondrial division in mammalian cells. Given the large number of regulatory mechanisms, it is possible that variations on these two pathways, or completely independent pathways, remain to be found.

Nature **593**, 346-347 (2021)

doi: <https://doi.org/10.1038/d41586-021-01173-x>

## References

1. 1.

Lavoisier, A. L. in *Œuvres de Lavoisier* Vol. II, 691–692 (Imprimerie impériale, 1862).

2. 2.

Kleele, T. *et al.* Nature **593**, 435–439 (2021).

3. 3.

Kennedy, E. P. & Lehninger, A. L. *J. Biol. Chem.* **179**, 957–972 (1949).

4. 4.

Ernster, L. & Schatz, G. *J. Cell Biol.* **91**, 227s–255s (1981).

5. 5.

Martinez-Vicente, M. *Front. Mol. Neurosci.* **10**, 64 (2017).

6. 6.

Kraus, F., Roy, K., Pucadyil, T. J. & Ryan, M. T. *Nature* **590**, 47–66 (2021).

7. 7.

Abrisch, R. G., Gumbin, S. C., Wisniewski, B. T., Lackner, L. L. & Voeltz, G. K. *J. Cell Biol.* **219**, e201911122 (2020).

8. 8.

Chakrabarti, R. *et al.* *J. Cell Biol.* **217**, 251–268 (2018).

9. 9.

Wolf, D. M. *et al.* *EMBO J.* **38**, e101056 (2019).

10. 10.

Peng, W., Wong, Y. C. & Krainc, D. *Proc. Natl Acad. Sci. USA* **117**, 19266–19275 (2020).

11. 11.

Simpson, C. L. *et al.* *Cell Rep.* **34**, 108689 (2021).

12. 12.

Yamano, K., Fogel, A. I., Wang, C., van der Bliek, A. M. & Youle, R. J. *eLife* **3**, e01612 (2014).

13. 13.

Yu, R., Jin, S. B., Lendahl, U., Nistér, M. & Zhao, J. *EMBO J.* **38**, e99748 (2019).

14. 14.

Cho, B. *et al. Nature Commun.* **8**, 15754 (2017).

## Jobs from Nature Careers

- - - [All jobs](#)
    - 
    - [Post-Doc - Numerical and microfabrication development of 4D metamaterials for mechanical, acoustic and photonics applications](#)  
[Université Bourgogne Franche-Comté \(UBFC\)](#)  
[BESANCON, France](#)  
[JOB POST](#)
- [Postdoctoral Position - Ecological Modeler](#)  
[The University of British Columbia \(UBC\)](#)  
[Vancouver, Canada](#)

## JOB POST

- [Postdoctoral Fellow | Zandstra Stem Cell Bioengineering Lab](#)

[The University of British Columbia \(UBC\)](#)

[Vancouver, Canada](#)

## JOB POST

- [Masterthesis / internship \(m/f/x\)](#)

[Helmholtz Centre for Environmental Research \(UFZ\)](#)

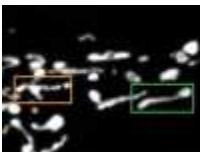
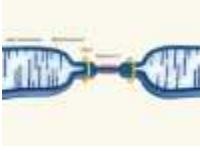
[Leipzig, Germany](#)

## JOB POST

[Download PDF](#)

[Download PDF](#)

## Related Articles

-  [Read the paper: Distinct fission signatures predict mitochondrial degradation or biogenesis](#)
-  [Double agents for mitochondrial division](#)



- **Mitochondria are mixed during cell division**
- **See all News & Views**

---

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-021-01173-x>

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

## NEWS AND VIEWS

28 April 2021

# Conversations, and how we end them

How we feel about the duration of our conversations has rarely been studied. New research has asked people about the lengths of their conversations, and whether they end when they want them to.

- [Elizabeth Stokoe](#) 0

### 1. [Elizabeth Stokoe](#)

1. Elizabeth Stokoe is in the School of Social Sciences and Humanities, Loughborough University, Loughborough LE11 3TU, UK.

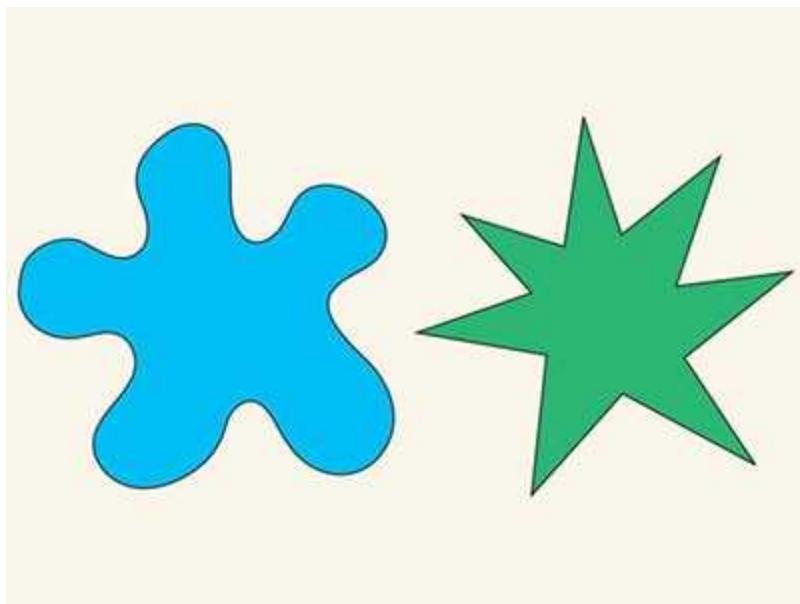
[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

Conversation has been described<sup>1</sup> as “the primordial site of human sociality”. We all have a lifetime’s experience to draw on if asked how it works, or when we reflect on the conversations we have participated in. But because conversation is something that we know tacitly how to do, scientific attempts to understand it are often relegated to the ‘soggy’ end of social psychology. Conversation certainly differs from other subjects of scientific scrutiny. For instance, black holes do not exist to be understood by people, whereas conversation exists only to be understood by people and to help us understand each other. Writing in *Proceedings of the National Academy of Sciences*, Mastroianni *et al.*<sup>2</sup> report how they have taken up the challenge of researching conversation scientifically.

The authors focused on the question of whether conversations end when people want them to, and gathered data from two studies. In the first one, individuals (806 in total) taking part in an online survey were asked to recall the most recent conversation they had in person, report its duration and indicate whether it ended when they wanted it to. If they indicated that the conversation didn't end when they wanted, they were asked to estimate how much longer or shorter they would have liked it to have been.

Participants were also asked how they thought the person they were speaking to might have answered the same questions. These conversations were mostly between people who were familiar to each other; 88% were between those who had known each other for at least a year, and 84% of the participants spoke to the person in question at least a few times each week.



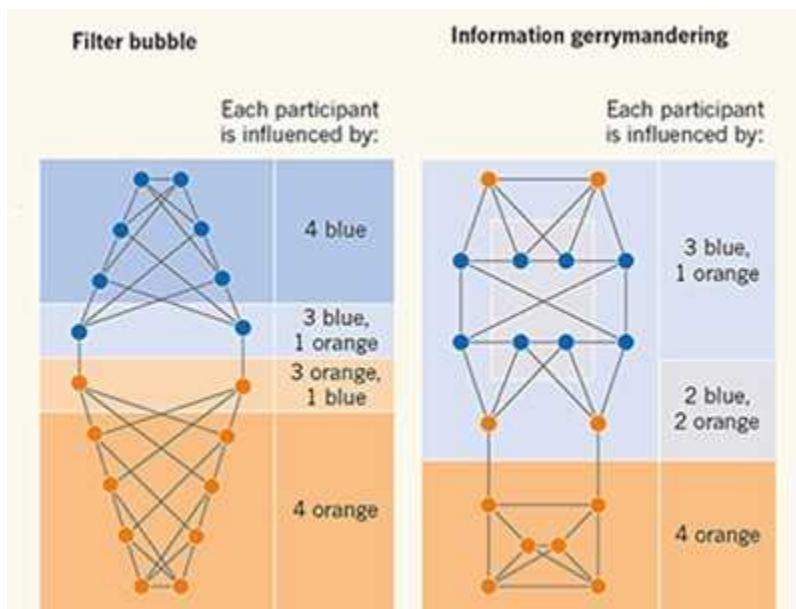
### Sound and meaning in the world's languages

The authors explain that their second study was designed, in part, to deal with the limitations of the first, such as relying on participant recall of an event and accessing only one party's view of the interaction. In the second study, the authors brought 366 previously unacquainted participants from university study pools into the laboratory for a one-hour experiment. Participants were paired up to have a conversation about whatever they wanted, for any duration, up to a maximum of 45 minutes. This was

followed by another task until the 60 minutes were up. The participants answered the same questions as those used in the online survey.

The authors concluded that the conversations evaluated in both studies almost never ended when both individuals wanted them to, and rarely ended when even one person wanted them to. Interaction with intimates or strangers made no difference to this mismatch. Participants were unaware of when their partner wanted to end the interaction, or that their own perceptions were so different from their partner's. For example, in laboratory conversations, there was a 46% discrepancy between the actual and desired duration. Interestingly, the authors excluded 57 of the pairs because they spoke for the full 45 minutes and did not end their conversations.

The study's novelty is in its examination of how people feel about their conversations at this level of empirical scrutiny. It represents a clear advance in psychology, in getting closer to where the action of social life happens, especially in the second study. One future direction for research might be to record or analyse the laboratory conversations themselves, and to ask participants to use transcripts to inform their responses. Transcripts would help the participants to identify precisely the point at which they wanted the conversation to end, and help researchers to understand exactly what each party was doing at the time. Were they mid-story, repeating something or giving a minimal or an expansive response? People show how attuned they are to tiny nuances in social interaction even as it unfolds<sup>3</sup>, and transcripts might enable the authors to gain extra insights about their findings.



### Information gerrymandering in social networks skews collective decision-making

Some have commented that, despite psychology being a discipline associated with “professional people watchers”<sup>4</sup>, psychologists rarely investigate “where moment-to-moment behaviour naturally happens”<sup>4</sup>, or deploy “direct observation of actual behaviour”<sup>5</sup>. Indeed, Mastroianni *et al.* say that scientists know little about conversation: “how it starts, how it unfolds, or how it ends.” One possible direction for future research, therefore, is to combine laboratory studies of the kind conducted in the second study by Mastroianni and colleagues with investigations of naturally occurring talk.

For more than 50 years, the cumulative science of conversation analysis has examined audio and video recordings of anywhere from single cases to thousands of cases of conversation. One benefit of augmenting laboratory studies and surveys with such data and methods is to avoid the limitations of post-hoc survey methods as identified by Mastroianni *et al.*, as well as the limitations of laboratory settings. All conversations have a reason for occurring, whether mundane or dramatic. In the authors’ second study, the reason was to be a research participant, making the experimental setting itself the ‘invariant occasion’ for the conversations that happened<sup>6</sup>. We know, however, that people interact differently when they are in a

simulation or experiment compared with their behaviour in life ‘in the wild’, because the reason for the event and their stake in its outcome are different<sup>7,8</sup>.

The authors conclude that people cannot coordinate what each participant wants from a conversation, in terms of ending it at a mutually satisfactory time, according to their responses when asked later. This striking observation tells us something interesting about the difference between what happens inside a conversation and what people say about it afterwards. As Mastroianni *et al.* point out, analysis has shown that conversations have ‘closing rituals’, which are systematically coordinated. In other words, a typical conversation does not usually end abruptly; it must be brought to a close<sup>7</sup>. Endings take shape through highly routine practices, such as making arrangements (“So let’s sort out what time on Monday”), or re-stating the reason for the conversation (“Well, I just wanted to see how you were doing”), combined with a ‘terminal exchange’<sup>9</sup> such as:

**A:** Okay

**B:** Okay

**A:** Bye bye

**B:** Bye

Endings are so systematic and recognizable that it can be easy to locate in transcripts the place at which someone wants the conversation to be over, whether by giving a delayed or minimal response or by saying something that indicates they are moving to draw the conversation to a close.

For example, in a study of individuals calling their doctors<sup>10</sup>, receptionists often initiated the end of a call before the caller was ready. In the following example from that study, the caller ‘wants’ the call to continue after the receptionist has started to end it with “Okay then”, followed by “Thank you”. The square brackets indicate when both spoke at the same time. The final four lines are a classic ‘terminal exchange’.

**Receptionist:** Okay then,

*(pause of 0.5 seconds)*

**Caller:** [ So it's th- ]

**Receptionist:** [Thank you, ]

*(pause of 0.5 seconds)*

**Caller:** That's the sixteenth?

**Receptionist:** The sixteenth, [ at ten pa]st eleven.

**Caller:** [Okay then.]

*(pause of 0.3 seconds)*

**Caller:** Ten past eleven, thank you.

**Receptionist:** Thank you,

*(pause of 0.2 seconds)*

**Caller:** T[hank you,]

**Receptionist:** [ Bye ]

**Caller:** Bye.

Furthermore, there was a correlation between surgeries in which the receptionists, rather than the callers, moved to initiate the ending of the call and lower patient-satisfaction scores with the surgery in general. Scrutiny of conversation transcripts reveals why this type of scenario creates dissatisfaction: in this case, the individual had to push past the receptionist's move to end the call to get their appointment confirmed.

When one party walks out or puts the phone down on another, we have the exact definition of mismatched desires regarding a conversation's ending. In the following call<sup>11</sup>, a salesperson 'cold calls' a company with the goal of selling printing systems (transcript simplified). Even when the call is

nearing its end, the client still uses components of the terminal exchange before hanging up.

**Client:** Well we're happy with, uhm, the people that we're currently using.

*(pause of 0.3 seconds)*

**Sales:** I'm sure you are, but I wanted to find out when the contract's up for review so then I can c[all maybe nearer the time]

**Client:** [Yeah no we're happy with wh]o we're currently using.

*(pause of 0.5 seconds)*

**Sales:** You don't know when the [contract's up for re]view?

**Client:** [Okay. Thank you.]

*(hangs up)*

Asking ‘when did you want the call to end?’ is the wrong question here, because the conversation is likely to be unwanted by the client in the first place, and, once in it, the parties involved want different outcomes.

Both the medical and the sales calls show that, and indicate how, individuals ‘want’ their conversations to end at different points. We can identify this in real settings in which we understand the authentic purpose of the conversation. It would therefore be interesting to apply Mastroianni and colleagues’ methods to the analysis of such transcripts and recordings, to ask individuals later, on reflection, to identify at what point they wanted to continue or end the conversations.

What about conversations between loved ones — such as those recollected in the online survey? In the following conversation<sup>12</sup> (transcript simplified) between Sue (not her real name), a young person with learning disabilities and in residential care, and her dad, Sue asks her dad to bring her extra pocket money when he visits. This is followed by the first turn that moves to close the conversation:

**Dad:** Right, well, I'm gonna get on now, I'll be there for about half past nine tomorrow morning.

But the conversation continues for a further 45 seconds before another pre-closing event occurs:

**Dad:** Right, well I'm going to go now, darlin'.

**Sue:** Yeah I've got to finish my cards off.

Only after three more pre-closings, including those expressing love (Dad: "Okay, lovey?" Sue: "yeah"; Dad: "I love you"; Sue: "love you"), do they bring the call to its end.

How do you show that you care about someone? Mastroianni *et al.* rightly point out that conversation is the "bread and butter" of our psychological and physical health, and this is clear to see in Dad and Sue's conversation. Staying longer in the conversation than external constraints allow (such as in a film scene in which people in a lift miss their floor to keep talking) is one way to do it. Closing rituals are so systematic that the conversational machinery allows us to see how the reopening of closings happen.

Mastroianni and colleagues' findings are compelling. Some media headlines about their study (see [go.nature.com/3sgIkup](http://go.nature.com/3sgIkup)), such as "only 2% of conversations end when we want them to", focused on the disconnect between the desired point for a conversation to end and its actual end. Although the headline news might be the scale of the disconnect, reducing conversations such as this chat between Dad and Sue to 'who wanted what' damages the empirical reality of their conversation and misses its purpose.

There are tremendous real-world benefits to analysing conversation with close scrutiny and rigour. For example, returning to the doctor's surgery, the same research<sup>10</sup> showed that when receptionists proactively confirmed an individual's appointment time and date, rather than doing so only in response to a request for confirmation, the conversation ended collaboratively. Moreover, proactive confirmation was associated with higher patient satisfaction, and the finding was used to train receptionists.

Do conversations end when people want them to? Mastroianni *et al.* conclude that the answer is almost certainly no. Asking people to report on their conversations has shown this clearly. Apart from situations such as in an argument, people generally do not say, “I want this conversation to end.” They might tell other individuals, “I was trapped in that conversation for hours”, or “I don’t want to talk to her”, but, in real conversation, people usually convey such things tacitly. This is why examining conversations, including using transcripts, is informative. It is clear, as Mastroianni *et al.* state, that “The more we learn about conversation — about how it begins and ends, runs and stalls, delights and disappoints — the better positioned we will be to maximize its benefits.”

Nature **593**, 347-349 (2021)

doi: <https://doi.org/10.1038/d41586-021-01074-z>

## References

1. 1.

Schegloff, E. A. *Social Psychol. Q.* **50**, 101–114 (1987).

2. 2.

Mastroianni, A. M., Gilbert, D. T., Cooney, G. & Wilson, T. D. *Proc. Natl Acad. Sci. USA* **118**, e2011809118 (2021).

3. 3.

Strivers, T. *Res. Lang. Social Interact.* **41**, 31–57 (2008).

4. 4.

Mehl, M. R. *Curr. Direct. Psychol. Sci.* **26**, 184–190 (2017).

5. 5.

Baumeister, R. F., Vohs, K. D. & Funder, D. C. *Perspect. Psychol. Sci.* **2**, 396–403 (2007).

6. 6.

Schegloff, E. A. *J. Narrative Life Hist.* **7**, 97–106 (1997).

7. 7.

Stokoe, E. *Res. Lang. Social Interact.* **46**, 165–185 (2013).

8. 8.

Stokoe, E., Sikveland, R. O., Albert, S., Hamann, M. & Housley, W. *Disc. Stud.* **22**, 87–109 (2020).

9. 9.

Schegloff, E. A. & Sacks, H. *Semiotica* **8**, 289–327 (1973).

10. 10.

Stokoe, E., Sikveland, R. O. & Symonds, J. *Br. J. Gen. Pract.* **66**, e779–e785 (2016).

11. 11.

Humă, B. & Stokoe, E. *Res. Lang. Social Interact.* **53**, 271–294 (2020).

12. 12.

Patterson, A. & Potter, J. *Br. J. Social Psychol.* **48**, 447–465 (2009).

## Jobs from Nature Careers

•

◦

- [All jobs](#)
  - 
  - [Post-Doc - Numerical and microfabrication development of 4D metamaterials for mechanical, acoustic and photonics applications](#)

[Université Bourgogne Franche-Comté \(UBFC\)](#)

[BESANCON, France](#)

[JOB POST](#)

- [Postdoctoral Position - Ecological Modeler](#)

[The University of British Columbia \(UBC\)](#)

[Vancouver, Canada](#)

[JOB POST](#)

- [Postdoctoral Fellow | Zandstra Stem Cell Bioengineering Lab](#)

[The University of British Columbia \(UBC\)](#)

[Vancouver, Canada](#)

[JOB POST](#)

- [Personalreferent\\*in \(d/m/w\), Gruppe Personaladministration](#)

[Helmholtz Centre for Heavy Ion Research GmbH \(GSI\)](#)

[Darmstadt, Germany](#)

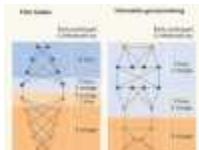
## JOB POST

[Download PDF](#)

[Download PDF](#)

## Related Articles

- [Read the paper: Do conversations end when people want them to?](#)



- [Information gerrymandering in social networks skews collective decision-making](#)



- [Sound and meaning in the world's languages](#)
- [See all News & Views](#)

---

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-021-01074-z>

## NEWS AND VIEWS

19 May 2021

# Iron and nickel vapours are present in most comets

The detection of iron and nickel vapours in a broad range of Solar System comets, and of nickel vapour in a comet from outside the Solar System, provides a glimpse into the organic chemistry of young planetary systems.

- [Dennis Bodewits](#) <sup>0</sup> &
- [Steven J. Bromley](#) <sup>1</sup>

### 1. [Dennis Bodewits](#)

1. Dennis Bodewits is in the Department of Physics, Auburn University, Auburn, Alabama 36832, USA.

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

### 2. [Steven J. Bromley](#)

1. Steven J. Bromley is in the Department of Physics, Auburn University, Auburn, Alabama 36832, USA.

[View author publications](#)

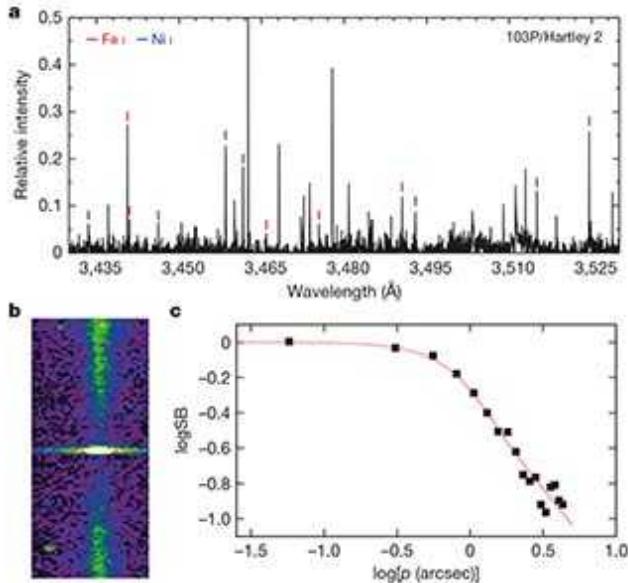
You can also search for this author in [PubMed](#) [Google Scholar](#)

Comets are agglomerates of dust and ice — leftovers from the era of planet formation. For most comets, their distance from the Sun keeps their temperature below a few hundred kelvin, which is still hot enough for water ice and other volatile compounds to sublime (be converted directly from

solid to gas). Comet nuclei are mostly obscured by a surrounding cloud of gas and dust called the coma. Therefore, knowledge of comet surfaces and their composition must be inferred from observations of the coma. Typical telescopic observations of cometary comae do not detect metals, because temperatures at comet surfaces are too low for these elements to sublime. However, two papers<sup>1,2</sup> in *Nature* this week report the discovery of metal atoms in cometary atmospheres, begging the question of where these atoms come from.

There have been several space missions to comets, including Rosetta, Deep Impact and Stardust. These missions have shown that comets are relatively small (typically, just a few kilometres in radius), and might be responsible for moving volatile materials around in the inner Solar System after the planets formed<sup>3</sup>. Such missions provided detailed studies of individual comets, but Earth-based observations have determined the chemical composition of larger numbers of these bodies<sup>4,5</sup>.

At optical wavelengths, the spectra of light emitted by comets coincidentally resemble those of flames. They have a broad, continuous part (caused, in flames, by hot soot; in comets, by dust that reflects sunlight), combined with the emission features of molecules and their fragments, such as hydroxyl (OH), cyanide (CN) and dicarbon (C<sub>2</sub>) groups. Until now, emission lines of metals — iron, nickel and other heavy elements — were thought to be absent from comet spectra. The detection of lone metal atoms in comets has been limited to specific situations, including sample-return missions (Stardust<sup>6</sup>) and bright, ‘sungrazing’ comets such as Ikeya–Seki, which plunged into the Sun<sup>7</sup>.



[Read the paper: Iron and nickel atoms in cometary atmospheres even far from the Sun](#)

Manfroid *et al.*<sup>1</sup> used atomic models to predict at which wavelengths, and how strongly, [iron and nickel emit light when illuminated by sunlight](#). They then identified dozens of emission lines of atomic iron and atomic nickel in the spectra of a broad sample of Solar System comets. The authors and their collaborators had observed these comets over the past two decades using the Very Large Telescope in Chile. The emission lines were hiding in plain sight, mingled among the typical (and plentiful) emission features of molecules in the coma.

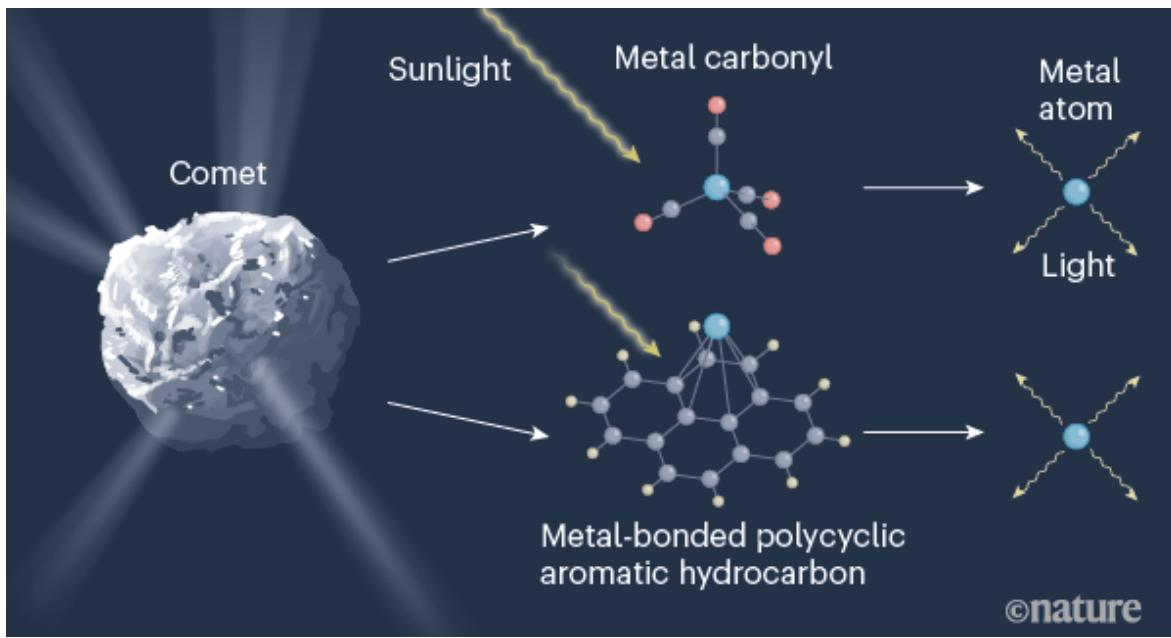
By looking at the spatial distribution of the light emitted by the iron and nickel atoms, Manfroid *et al.* calculated the rate of metal loss from the comets. The amount of iron and nickel released is surprisingly small — only about 1 gram per second, compared with the roughly 100 kilograms per second of water produced. This finding testifies to the remarkable sensitivity of the Very Large Telescope. Coincidentally, the amount of nickel produced per second is almost exactly the nickel content of a US five-cent coin, or nickel.

As for the source of the iron and nickel atoms, their spatial distribution suggests they are formed in the coma, close to the nucleus. A second clue to their origin lies in the comets sampled by Manfroid and colleagues. Despite

large differences in the mass-loss rate, distance to the Sun and chemical composition of individual comets, iron and nickel were found in all comets that were studied in detail. The authors estimated that temperatures reach only about 150 K at the most distant comet examined. This result suggests that the source of the iron and nickel atoms is much more volatile than the sulfides of these metals (which are found in cometary dust grains) or the pure metals.

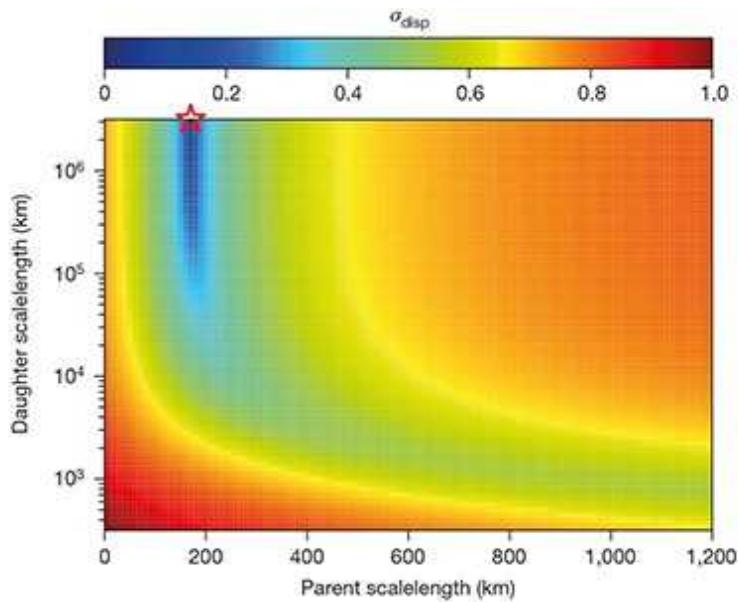
A third clue to the origin of these atoms is that the amount of nickel relative to iron is much higher in the studied comets than in the Sun, meteoroids and sungrazing comets. Moreover, the observed spectra lack emission lines of other metals, such as chromium and manganese, seen in the spectra of sungrazing comets. These observations rule out direct sublimation of metal particles, and indicate an intermediate chemical or physical process that determines the nickel-to-iron ratio in the coma.

Manfroid and colleagues provide several possible production scenarios. Under certain conditions, tiny dust grains in the coma could reach temperatures above 1,000 K and release the metals into the coma. Alternatively, the metals could initially be locked inside organometallic compounds, such as nickel or iron carbonyls, which consist of four or five carbon monoxide (CO) molecules bound to a nickel or iron atom (Fig. 1). Modelling by the authors suggests that these carbonyls can sublime at temperatures as low as 74 K, similar to carbon dioxide. However, the release mechanism is problematic: experiments indicate that the CO molecules would be stripped off the carbonyls sequentially, rather than the metal atoms being released directly<sup>8</sup>.



**Figure 1 | How comets could release metal atoms.** Manfroid *et al.*<sup>1</sup> and Guzik and Drahus<sup>2</sup> have detected the light emitted by metal atoms in the atmospheres of comets. The surfaces of these comets do not reach high enough temperatures to release such atoms directly. Instead, the source of the atoms might be organometallic compounds that are emitted from the comet's surface and then break up in the comet's atmosphere when irradiated by sunlight. These compounds could be metal carbonyl complexes, which consist of carbon monoxide molecules bound to a metal atom, or metal-bonded polycyclic aromatic hydrocarbons, which are sheets of carbon atoms bordered by hydrogen atoms and attached to a metal ion. Carbon, grey; oxygen, red; hydrogen, white.

An alternative source of the iron and nickel atoms could be metal-bonded polycyclic aromatic hydrocarbons, which are sheets of carbon atoms bordered by hydrogen atoms and attached to a metal ion (Fig. 1). Such compounds might also sublime and rapidly break up in the coma when exposed to the Sun's harsh ultraviolet light. Intriguingly, the abundance of these compounds in cometary atmospheres has been reported to be one for every one million water molecules<sup>9</sup> — similar to the relative abundance observed for the iron and nickel atoms.



[Read the paper: Gaseous atomic nickel in the coma of interstellar comet 2I/Borisov](#)

Independently of Manfroid *et al.*, Guzik and Drahus<sup>2</sup> detected [light emission from atomic nickel around the comet 2I/Borisov](#) using the Very Large Telescope in January 2020. The orbit of this comet indicated that it came from outside the Solar System, and, surprisingly, both its behaviour and composition of free radicals had much in common with those of regular comets. However, its high content of CO gas suggested that it probably formed under very different circumstances from those for Solar System comets. On the basis of this anomaly, it has been proposed that 2I/Borisov is a fragment of a larger, Pluto-like object<sup>10</sup>, or formed around a star smaller and colder than the Sun — such as an M dwarf, the most common type of star in the Galaxy<sup>11</sup>.

Guzik and Drahus report a nickel abundance for 2I/Borisov that is similar to that found by Manfroid *et al.* for Solar System comets. Given the unknown chemical origins and physical history of 2I/Borisov, this similarity is striking. If we can unravel the origin of iron and nickel in regular comets and this interstellar object, we might uncover a story of organic chemistry between shared different planetary systems.

doi: <https://doi.org/10.1038/d41586-021-01265-8>

## References

1. 1.  
Manfroid, J., Hutsemékers, D. & Jehin, E. *Nature* **593**, 372–374 (2021).
2. 2.  
Guzik, P. & Drahus, M. *Nature* **593**, 375–378 (2021).
3. 3.  
Rubin, M., Bekaert, D. V., Broadley, M. W., Drozdovskaya, M. N. & Wampfler, S. F. *ACS Earth Space Chem.* **3**, 1792–1811 (2019).
4. 4.  
A'Hearn, M. F., Millis, R. C., Schleicher, D. G., Osip, D. J. & Birch, P. V. *Icarus* **118**, 223–270 (1995).
5. 5.  
Dello Russo, N., Kawakita, H., Vervack, R. J. Jr & Weaver, H. A. *Icarus* **278**, 301–332 (2016).
6. 6.  
Brownlee, D. *Annu. Rev. Earth Planet. Sci.* **42**, 179–205 (2014).
7. 7.  
Preston, G. W. *Astrophys. J.* **147**, 718–742 (1967).
8. 8.  
Leadbeater, N. *Coord. Chem. Rev.* **188**, 35–70 (1999).

9. 9.

Klotz, A. *et al.* *Planet. Space Sci.* **44**, 957–965 (1996).

10. 10.

Cordiner, M. A. *et al.* *Nature Astron.* **4**, 861–866 (2020).

11. 11.

Bodewits, D. *et al.* *Nature Astron.* **4**, 867–871 (2020).

## Competing Interests

The authors declare no competing interests.

## Jobs from Nature Careers

- - - [All jobs](#)
    - - [\*\*Post-Doc - Numerical and microfabrication development of 4D metamaterials for mechanical, acoustic and photonics applications\*\*](#)  
[Université Bourgogne Franche-Comté \(UBFC\)](#)  
[BESANCON, France](#)  
[JOB POST](#)
      - [\*\*Postdoctoral Position - Ecological Modeler\*\*](#)  
[The University of British Columbia \(UBC\)](#)

[Vancouver, Canada](#)

[JOB POST](#)

- [Postdoctoral Fellow | Zandstra Stem Cell Bioengineering Lab](#)

[The University of British Columbia \(UBC\)](#)

[Vancouver, Canada](#)

[JOB POST](#)

- [Personalreferent\\*in \(d/m/w\), Gruppe Personaladministration](#)

[Helmholtz Centre for Heavy Ion Research GmbH \(GSI\)](#)

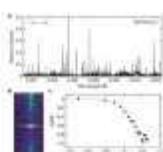
[Darmstadt, Germany](#)

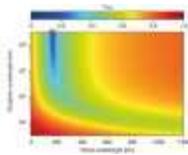
[JOB POST](#)

[Download PDF](#)

[Download PDF](#)

## Related Articles

-  [Read the paper: Iron and nickel atoms in cometary atmospheres even far from the Sun](#)



- [Read the paper: Gaseous atomic nickel in the coma of interstellar comet 2I/Borisov](#)



- [Eye of a skull reveals details of cometary materials](#)



- [Cometary spin-down](#)

- [See all News & Views](#)

---

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-021-01265-8>

Unconventional viral gene expression mechanisms as therapeutic targets

[Download PDF](#)

- Review Article
- [Published: 19 May 2021](#)

# Unconventional viral gene expression mechanisms as therapeutic targets

- [Jessica Sook Yuin Ho<sup>1,na1</sup>](#),
- [Zeyu Zhu<sup>1,na1</sup>](#) &
- [Ivan Marazzi](#)    [ORCID: orcid.org/0000-0003-3376-0922<sup>1,2</sup>](#)

*Nature* volume 593, pages 362–371 (2021) [Cite this article](#)

- 3527 Accesses
- 7 Altmetric
- [Metrics details](#)

## Subjects

- [Infection](#)
- [Pathogens](#)
- [Viral genetics](#)
- [Virology](#)

## Abstract

Unlike the human genome that comprises mostly noncoding and regulatory sequences, viruses have evolved under the constraints of maintaining a small genome size while expanding the efficiency of their coding and regulatory sequences. As a result, viruses use strategies of transcription and translation in which one or more of the steps in the conventional gene–protein production line are altered. These alternative strategies of viral gene expression (also known as gene recoding) can be uniquely brought about by dedicated viral enzymes or by co-opting host factors (known as host dependencies). Targeting these unique enzymatic activities and host factors exposes vulnerabilities of

a virus and provides a paradigm for the design of novel antiviral therapies. In this Review, we describe the types and mechanisms of unconventional gene and protein expression in viruses, and provide a perspective on how future basic mechanistic work could inform translational efforts that are aimed at viral eradication.

[Download PDF](#)

## Main

Expression of a gene in the human genome is a multistep and heavily regulated process that resembles a production line. Protein-coding genes are transcribed almost exclusively by RNA polymerase II (RNAPII). During transcription, quality-control checkpoints are implemented to ensure that a gene is properly recognized and transcribed. A number of factors (epigenetic enzymes, chromatin remodelers, transcription factors and activators–coactivators) ensure gene recognition and RNAPII progression on the genic template. The progression of RNAPII—which includes RNAPII initiation, pause–release, elongation and the termination of transcription—occurs in sync with co-transcriptional events (that is, 5' capping, splicing and polyadenylation). The end result of gene transcription and RNA processing is the generation of a mature RNA, in which coding exons are fused in a linear order that depends on the isoform of the gene. Mature mRNA is subsequently exported from nucleus into the cytoplasm, where it is directed to ribosomes for translation. The canonical model of translation initiation starts with recognition of the 7-methylguanylate cap on the 5' end of most eukaryotic mRNA by the initiation factor eIF4, which recruits a pre-initiation complex that comprises the 40S ribosomal subunit and several eukaryotic initiation factors (eIF3, eIF1, eIF1A and the ternary complex eIF2–GTP–Met-tRNA<sub>i</sub><sup>Met</sup>). This complex then scans continuously from the 5' to the 3' end for the first initiation codon in an optimal context (the RCCAUGG Kozak sequence, in which R stands for purine)<sup>1</sup>. Once the start codon of a gene is read by the initiator tRNA<sup>Met</sup>, translation progresses and ends when a stop codon in the mRNA (UAA, UAG or UGA) is recognized by release factors. Depending on the subcellular localization of a given protein, co- and post-translational events might take place to sort proteins to their destinations. In brief, this is the conventional eukaryotic production line through which a gene makes a protein ready to be used in the cell.

To overcome their small genomes and increase their coding capacity, viruses have evolved to co-opt the transcriptional, epigenetic and translational mechanisms of the infected host cell. To generate protein diversity, viruses can adopt the existing mechanisms of the host (for example, alternative splicing) or use unique strategies. Here we describe the diverse ways by which viral genomes give rise to genes and proteins that deviate from the canonical framework of human genes, restricting our analyses to eukaryotes and their viruses.

## Small-genome solutions to big problems

A main strategy to increase the number of coded proteins from a small genome is the use of overlapping or overprinted genes. Nucleic acid sequences can simultaneously encode two or more proteins in alternative reading frames (ARFs). To synthesize these proteins, unconventional transcriptional ('copying') or translational ('reading') events need to take place (Fig. 1). Although a comprehensive characterization of gene overprinting in large mammalian genomes is lacking, estimates on the basis of simulating codon use<sup>2</sup> or ribosome footprinting<sup>3</sup> suggest that only 1% of human genes are overprinted. By contrast, gene overlapping is very common among viruses. Despite differences in the size and structure of viral genomes, 53% of sequenced viral genomes containing at least one pair of genes that overlap for more than 50 nucleotides<sup>4</sup>. Proteins that originate by overprinting often encode accessory proteins that feature short sequences, and can provide a selective advantage for viruses<sup>5,6,7</sup>. Many overlapping genes are fixed in viral genomes because of their functions as host antagonists, such as those that affect the interferon response of the host<sup>8,9</sup>, suppress RNA interference<sup>10</sup>, and induce apoptosis of host cells<sup>11</sup>. In addition, as a mutation in an overlapping genomic region affects both the canonical and the overprinted genes, overlapping genes may also serve as a safety mechanism that protects the virus from deleterious mutations. However, because proteins that are encoded by gene overprinting are often enriched in disordered regions and show a tendency to have no known homologues<sup>12,13</sup>, many overprinting viral proteins are poorly characterized.

**Fig. 1: The host and virus adopt different strategies for gene expression as a result of differences in genome size.**



Left, in organisms with a large genome, expression of a cellular gene typically follows a linear pathway that leads to the synthesis of the respective canonical protein product. Right, viruses, which are confined by their much-smaller genome sizes, use unconventional pathways that mostly involve transcription-level (decoding multiple messages (mRNA)) or translation-level regulation to generate several protein products from a single locus.

[Full size image](#)

Another challenge that is inherent to a small genome is a lack of regulatory space for maintaining the correct stoichiometry and temporality of the expression of overprinted proteins. To overcome these limitations, viruses use several methods that include (1) intrinsic *cis* and *trans* regulation of polymerase and other enzymatic activities and (2) a codependency on host functions. We summarize the most relevant strategies used by viruses for expanding the coding and regulatory potentials of their overlapping genes, focusing mostly on viruses that are human pathogens and that represent current and future threats.

## Expression of overlapping genes

### Copying multiple messages

One set of strategies used by viruses to increase the efficiency of their small genomes involves transcriptional mechanisms that generate several mRNAs from overprinted coding sequences.

### Transcriptional slippage

Transcriptional slippage is a process in which several overlapping transcripts are generated from the same gene via viral RNA polymerase stuttering, which results in the incorporation (and, occasionally, the deletion) of one or more nucleotides in the transcript (Fig. 2a). Sequences that are prone to transcriptional slippage include homopolymeric A/T tracts, the U6A motif in human immunodeficiency virus (HIV)<sup>14</sup>, and the UC-rich slippery sequence in the paramyxoviruses<sup>15</sup>. The efficiency of transcriptional slippage is regulated by the stability and length of the nascent RNA relative to the template RNA, as well as by the structure of RNA-dependent RNA polymerase (RdRp)<sup>15</sup>. Owing to frameshift upon the insertion of nucleotides, the translation of overlapping transcripts typically results in proteins with a common N-terminus, but different C termini. Aside from using transcriptional slippage to generate mRNAs in different reading frames, some virus also use it to polyadenylate their mRNAs<sup>16</sup>.

**Fig. 2: Small-genome solutions to expanding coding potential.**

 [figure2](#)

**a**, Polymerase frameshifting, in which backward or forward slippage of RNA polymerase (pol) results in nucleotide insertions or deletions, and generates a heterogeneous population of viral mRNAs. **b**, PRFs lead to the synthesis of viral proteins from several reading frames. **c**, Leaky scanning, in which the ribosome scans through and skips an AUG start codon that is typically located in a less-optimal sequence context, and initiates at a downstream start codon. **d**, Generation of noncanonical sites of translation initiation through upstream ORFs or non-AUG start codons. In start-snatching, an upstream AUG start codon is obtained via cap-snatching of host RNA (which enables the translation of novel proteins on the basis of both host and viral genetic information).

[Full size image](#)

Transcriptional slippage was first identified in the synthesis of V proteins from the phosphoprotein (*P*) gene in *Parainfluenza virus 5* (previously known as *Simian virus 5*)<sup>17</sup>, and has subsequently been observed in other pathogenic RNA viruses: mostly of members of *Mononegavirales*, including viruses in the *Paramyxoviridae* (such as Sendai virus) and *Filoviridae* (such as ebolavirus). Positive-strand viruses in the *Potyviridae*<sup>18</sup> and *Flaviviridae*<sup>19</sup> families have also been described as using this mechanism. In paramyxoviruses, transcriptional slippage can occur when RdRp encounters a ‘slippery’ sequence of 3'-UUUUUUCCC-5' in the *P* gene and stutters at the underlined cytidine<sup>15</sup>. The polymerase then backtracks and realigns the newly synthesized mRNA with the template by non-destabilizing G:U base-pairing, which results in G insertions. The possible number of G insertions is limited to six by a sequence that contains adenosine that is located immediately upstream of the slippery site (as A:A base-pairing is not tolerated)<sup>20</sup>. In Sendai virus, at least three distinct mRNAs of the *P* gene are produced by transcriptional slippage. The unedited mRNA encodes P protein, which is a component of RdRp that regulates transcriptional fidelity and limits antiviral responses<sup>21,22</sup>. mRNA with +1 G or +2 G insertions code for two accessory proteins (V and W, respectively), both of which regulate viral replication kinetics and the activation of host responses<sup>23,24</sup>. Additionally, the unique hexameric genome-packaging rule of paramyxovirus might regulate the efficiency of mRNA editing mediated by transcriptional slippage in this virus<sup>20,25</sup>, as it has been shown that mRNA editing is at its most extensive when the cytidine at which the RdRp stutters is in position 2 or 5 in a hexamer, which suggests that N proteins might remain in close proximity to RdRp during transcription<sup>26</sup>. Further examples of transcriptional slippage occur in ebolaviruses and Marburg viruses<sup>27</sup>, both of which belong to the *Filoviridae* family. In ebolavirus, transcriptional slippage occurs at a 30% frequency on a stretch of seven uridines in the glycoprotein (*GP*) gene and results in the insertion of one or two additional adenines in the mRNA<sup>28,29,30,31</sup>. The unedited transcript translates into a nonstructural and secreted glycoprotein<sup>28</sup>, and the +1 A and +2 A shifts result in an extended glycoprotein that bears a transmembrane domain and a small soluble glycoprotein, respectively<sup>28</sup>. More recently, deep mRNA sequencing has revealed other possible polyuridine transcriptional slippage sites in the *GP*, *NP*, *VP30* and *L* mRNAs of ebolavirus<sup>27</sup>, which suggests that there may be more uncharacterized polypeptide species expressed than has previously been believed.

## RNA splicing

RNA splicing is a commonly used and tightly regulated eukaryotic mechanism of generating distinct mature transcripts from a single gene, and has also been exploited by several families of viruses that replicate in the host nucleus, such as members of the *Adenoviridae* and *Parvoviridae* (DNA viruses), retroviruses, and members of the *Bornaviridae* and *Orthomyxoviridae* (RNA viruses). However, because of the more

compact nature of viral genomes, splicing in viruses—unlike in humans—often serves to express overprinted genes.

In the segmented RNA genome of influenza A viruses (IAV), splicing occurs in viral segments 8 (which encodes the *NS* gene), 7 (which encodes the *M* gene) and 2 (which encodes the *PB1* gene). Depending on the viral strain, up to three or four unique mRNAs can be generated from segments 8 and 7, respectively. The noncanonical proteins that are produced by splicing are involved in important functions, such as the nuclear export of viral RNA and host adaptation<sup>32,33</sup>. Importantly, the splicing of segments 7 and 8 is regulated by an array of viral and host factors that includes *trans* regulators of splicing, such as NS1-BP, HNRNPK<sup>34</sup>, SRSF1 (also known as SF2/ASF)<sup>35</sup>, SRSF3<sup>36</sup> and protein kinase CLK1<sup>36</sup>. Finally, *cis*-regulatory RNA secondary structures at the 3' splice site of segment 7 have been suggested to be potential regulators of splicing efficiency in IAV<sup>37,38</sup>, and a determinant of host tropism<sup>37</sup>.

Circular RNA is a relatively stable and exonuclease-resistant RNA that is produced by backsplicing, and has recently been identified<sup>39</sup> across many viruses—including members of the gammaherpesvirus family (Epstein–Barr virus and Kaposi sarcoma virus) and the oncogenic human papillomaviruses. The functions of circular RNA in viruses are largely unknown, but a recent study has shown that knockdown of the E7 circular RNA produced by human papillomavirus 16 using short hairpin RNA inhibits oncogenic transformation of infected cells<sup>40</sup>.

## Reading multiple messages

Other mechanisms used by viruses to expand the set of proteins expressed from their small genomes include those that act at the level of mRNA translation, which allow for the expression of multiple overprinted proteins from one mRNA.

## Programmed ribosome frameshifting

Programmed ribosomal frameshifts (PRFs) (Fig. 2b) occur when elongating ribosomes slip by one base upstream (5', known as a -1 PRF) or downstream (3', known as a +1 PRF), thus shifting the ribosomal reading frame. PRFs allow for the expression of overprinted proteins from the same mRNA and can also serve to regulate the stoichiometry of viral proteins. There are two prerequisites for a -1 PRF: (1) a slippery site with the sequence motif XXXYYY<sub>Z</sub> (in which X is any three identical nucleotides, Y represents U or A, and Z is A, C or U (although with some exceptions, such as GGU); as has previously been reviewed in detail<sup>41,42</sup>) and (2) a downstream pseudoknot structure that comprises two stems and a connecting loop as a stimulatory

element for ribosomal pausing at the slippery site<sup>43,44</sup>. In +1 PRFs, ribosome pausing is also directed by the presence of rare or ‘hungry’ codons at the slippery site, which shifts the ribosomal A site onto a more abundant codon to resume elongation.

Much of our early understanding of –1 PRFs came from studies of the Rous sarcoma virus<sup>7</sup> and HIV-1<sup>45</sup>, in both of which the structural protein precursor (Gag) and the enzyme precursor (Pol) are translated from the same viral mRNA. Gag is produced through conventional translation. A –1 PRF midway through Gag synthesis occurs in 2–10% of translating ribosomes and results in a fusion protein that is known as Gag–Pol, which is later cleaved by viral proteases to generate full-length Pol<sup>5,46,47</sup>. PRFs also have an important role in members of the *Coronaviridae* (for example, severe acute respiratory syndrome coronavirus (SARS-CoV), severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) and Middle Eastern respiratory syndrome coronavirus) and *Flaviviridae* (for example, West Nile virus)<sup>48,49</sup>. In the *Coronaviridae*, the replicase gene is organized into two partially overlapping open reading frames (ORFs) known as ORF1a and ORF1ab that encode polyprotein 1a and the fused polyprotein 1a–1b, respectively, the latter of which is generated by a –1 PRF. This frameshift event occurs at a frequency of 14–27%<sup>50</sup>, and has been suggested as a mechanism that maintains the ratio of ORF1a to ORF1ab<sup>51</sup>. Unlike members of the *Retroviridae*, SARS-CoV contains an atypical three-stem pseudoknot and an additional, structurally conserved attenuator sequence that is 5' of the PRF signal<sup>50,51,52</sup>, which has been shown to control the frequency of –1 PRFs in coronaviruses<sup>51,52</sup>. Notably, lowering the efficiency of frameshifts markedly reduces viral replication and infectivity<sup>6,51,53,54,55</sup>, which underscores the importance of the –1 PRF for these viruses. Importantly, host factors have been identified that interfere with virus PRFs. For instance, the human protein C19Orf66—first identified for its inhibitory effect on the replication of dengue virus<sup>56</sup>—has been shown to inhibit –1 PRFs in Gag–Pol synthesis<sup>57</sup>. C19Orf66 has further been shown to exhibit broad-spectrum activity in blocking PRFs in HIV-2, Rous sarcoma virus, human T lymphotropic virus and mouse mammary tumour virus<sup>57</sup>. Whether C19Orf66 functions only by limiting PRFs requires investigation, but targeting PRF factors could provide a selective and powerful antiviral strategy.

## Leaky scanning

In ribosomal leaky scanning, the ribosome skips a translation initiation site (especially if this site is located in the context of a weak Kozak sequence) and initiates at a downstream one (Fig. 2c). Many viruses—including retroviruses<sup>58</sup>, paramyxoviruses<sup>59</sup>, papillomaviruses<sup>60</sup> and bunyaviruses—adopt leaky scanning to express several proteins from one transcript<sup>61,62</sup>. In pandemic strains of HIV, a bicistronic mRNA transcript encodes a conserved upstream, small 81-amino-acid

protein known as Vpu, which confers a fitness advantage by degrading the CD4 viral receptor and enhancing virion release<sup>58,63,64,65,66</sup>. The bypassing of the Vpu start codon leads to initiation on a downstream start codon, which results in the synthesis of the viral envelope protein<sup>58</sup>. In the segmented RNA genome of IAV, leaky scanning can generate four proteins in addition to the canonical protein that is encoded by segment 2<sup>67</sup>. For example, a downstream AUG leads to the synthesis of PB1-F2, a protein that localizes to mitochondria and elicits a pro-inflammatory and pro-apoptotic effect on host cells<sup>11,68,69,70</sup>.

## Translation of upstream ORFs

Although viruses have a relatively short 5' untranslated region, an increasing body of evidence suggests that upstream ORFs that are led by upstream start codons (AUGs) can be translated (Fig. 2d). Upstream translation has widely been observed in DNA viruses and positive- and negative-sense RNA viruses, as well as in mammalian genomes<sup>71,72,73,74,75,76,77,78,79,80,81</sup>. Upstream ORFs in viruses have been suggested to have two major functional consequences. First, and similar to mammalian upstream ORFs<sup>78,79,80,81,82</sup>, many viral upstream ORFs suppress the translation of the downstream canonical ORF. For instance, in ebolavirus, an upstream ORF of the *L* gene (which is important for replication and RNA capping) suppresses the translation of the *L* ORF under normal conditions and enhances it under stress conditions<sup>75</sup>. This bimodal regulation fine tunes the synthesis of L protein and helps to maintain optimal polymerase activity<sup>75</sup>. Similarly, upstream ORFs can regulate the expression of viral proteins in coronaviruses (such as murine hepatitis virus and bovine coronaviruses) and in several DNA viruses (such as hepatitis B virus and human cytomegalovirus)<sup>72,74,75,76,77</sup>. Second, the products of upstream ORFs can be involved in regulating virulence and tropism. In the monopartite genome of enteroviruses, a highly conserved upstream ORF partially overprints the canonical polyprotein ORF<sup>71</sup> and encodes a putative transmembrane protein that facilitates viral release and invasion of echovirus 7 in human gut epithelial cells<sup>71</sup>.

## Initiation of translation from non-AUG codons

The translation of many virus genes has been shown to initiate on noncanonical start codons that are typically found upstream of the canonical AUG codon<sup>81</sup> (Fig. 2d). These noncanonical start codons fall mainly into two categories. First, a near-cognate start codon that normally varies by one nucleotide from AUG can be recognized by the initiator tRNA<sub>i</sub><sup>Met</sup>, which occurs at the P-site of the ribosome. For instance, the polycistronic P/C mRNA of Sendai virus and parainfluenza virus type 1 encodes five proteins (P, C, C', Y1 and Y2) from overlapping ORFs. The C' protein is generated by

the efficient initiation of translation from an upstream non-AUG codon (ACG for Sendai virus and GUG for parainfluenza virus type 1), which has a N' extension compared to the C protein<sup>83,84</sup>. Similar uses of non-AUG start codons (most frequently CUG, and sometimes GUG) have been identified in viruses that infect a wide range of hosts, including murine leukaemia virus<sup>85</sup>, human T cell lymphotropic virus type 1<sup>86</sup>, influenza virus<sup>87</sup>, soil-borne wheat mosaic virus<sup>88</sup> and equine infectious anaemia virus<sup>89</sup>. Second, the non-AUG start codon can be recognized by a non-methionine tRNA. In this case, the initiator tRNA<sub>i</sub><sup>Met</sup> is not required and translation initiates in the A site. This leads to proteins that start with non-methionine amino acids, which have mainly been identified in insect viruses<sup>90,91</sup>.

### Start-snatching to generate hybrid proteins

Translation in eukaryotic cells requires the recognition of the 5' methyl-7-guanosine (m<sup>7</sup>G) cap on mRNA. Segmented negative-sense RNA viruses in the order *Bunyavirales* and the families *Orthomyxoviridae* (for example, IAV) and *Arenaviridae* (for example, Lassa virus) do not encode capping enzymes, but instead rely on a process known as ‘cap-snatching’ to access cap-dependent translation. In this process, viral polymerase binds to the m<sup>7</sup>G cap of host RNA and cleaves off a short stretch (7–20 nucleotides in the case of IAV and about 7 nucleotides for Lassa virus) of host capped-RNA<sup>92,93</sup>. These host-derived fragments are then used as a primer to initiate the transcription of viral mRNAs<sup>94</sup>. As a consequence, mRNAs of segmented negative-sense RNA viruses exist as genetic hybrids, in which 5' sequence heterogeneity is provided by snatched host-derived sequences<sup>92,95,96,97</sup>.

Instead of merely providing a m<sup>7</sup>G cap, cap-snatched host sequences that bear AUGs also allow segmented negative-sense RNA viruses to express cryptic ORFs within their 5' untranslated regions (known as upstream viral ORFs). This process has been termed ‘start-snatching’ (Fig. 2d). During IAV infection, about 12% of host-derived cap-snatched sequences bear AUG start codons that confer translation. Depending on the reading frame of the host-derived AUG with respect to the viral RNA, these codons initiate the synthesis of either host–virus chimeric N-terminally extended viral proteins or novel polypeptides (up to 80 amino acids in length) that are overprinted with the major viral ORF<sup>98</sup>. Start-snatching and the genesis of upstream viral ORFs may be a way for segmented negative-sense RNA viruses to sample evolutionary space before gene functionalization. A recent study has shown that some strains of IAV have evolved to encode an AUG start codon in the untranslated region of the nucleoprotein segment. Expression of this N-terminally extended nucleoprotein increases viral virulence<sup>99</sup>.

### Additional mechanisms

Genome compaction in viruses has driven additional mechanisms that do not rely on genic overprinting to express several proteins from a single locus, which have previously been reviewed<sup>81</sup> and are summarized in Box 1.

## Box 1 Nonoverlapping gene expression in viruses

Several proteins can be generated without overprinting from one coding sequence through an array of strategies that are used pervasively by viruses.

### IRESs

IRESs (typically involving stem-loop and pseudoknot structures located upstream of a coding sequence) are widely adopted by viruses to circumvent cap dependency during translation. First identified in poliovirus and encephalomyocarditis viruses<sup>156,157</sup>, IRESs have frequently been discovered in viruses with uncapped positive-strand RNA genomes (mostly picornaviruses), a few DNA viruses<sup>158,159</sup> and in mammalian genomes<sup>160</sup>. Although IRESs are functionally similar, there is no consensus sequence or structure for them. Mechanistically, IRESs can be divided into two types: a common type that recruits ribosome via binding to eIFs and other RNA-binding proteins<sup>161,162</sup>, and a simpler type that directly recruits ribosome without eIFs (which has mainly been found in cricket paralysis virus)<sup>163,164</sup>.

### Ribosomal shunting

In ribosomal shunting (an alternative mechanism for the cap-dependent initiation of translation initiation), the 40S ribosome subunit bypasses the scanning of some segments of RNA by translocating to a downstream shunt acceptor site. This is typically enabled by stable hairpin structures that are formed by 5'-RNA leader sequences that block scanning. Shunting was first discovered in cauliflower mosaic virus<sup>165</sup>, and has been found in many plant pararetroviruses<sup>166</sup> and in animal viruses that include adenoviruses<sup>167</sup> and Sendai virus<sup>168</sup>.

### Translation reinitiation

In translation reinitiation, post-termination ribosomes remain on a polycistronic viral mRNA and reinitiate translation from a nearby start codon either upstream or downstream of the termination codon of the preceding ORF. Translation reinitiation was first discovered in the polycistronic subgenomic RNA of rabbit haemorrhagic disease virus, in which two ORFs overlap by 17 nucleotides and encode the major and minor capsid proteins and several nonstructural proteins<sup>169</sup>. Translation reinitiation also occurs in caliciviruses and negative-sense RNA viruses such as influenza B virus<sup>170</sup>, and human respiratory syncytial virus<sup>171,172</sup>.

## **Read-through translation**

In read-through translation, the ribosome reads a stop codon as a sense codon influenced by the stop codon context, which results in continued translation and protein products with extended C termini. Read-through translation has been observed on all three stop codons (most commonly on UGA) and can be mediated by either normal tRNAs or suppressor tRNAs. Read through is used extensively in many viruses, ranging from alphaviruses to mimivirus<sup>[173](#)</sup>. Although read through is relatively rare in mammalian genes (as has previously been reviewed<sup>[174](#)</sup>), it can result in insertions of selenocysteine<sup>[175,176](#)</sup>.

## **Stop–carry on**

In stop–carry on, the ribosome skips the formation of a peptidyl bond while reading two consecutive sense codons, which generates two peptides from one ORF in a stop-codon-independent manner. Stop–carry on was first identified in the peptidase 2A and 2B regions of the polyprotein of foot and mouth disease virus<sup>[177,178](#)</sup> (as has previously been reviewed<sup>[179](#)</sup>). The 2A peptide structurally hinders the binding of the last tRNA (tRNA-Pro) to the ribosomal A site but not to the release factors<sup>[180](#)</sup>, thus causing the skipping of peptidyl bond formation. Stop–carry on occurs mainly on 2A and 2A-like peptides, which are widely present and conserved in picornaviruses and other mammalian and insect viruses<sup>[179](#)</sup>.

## **Lessons for the development of therapeutic agents**

A fundamental principle that underlies the development of antiviral drugs is to evaluate the benefit (for example, infection suppression) versus the cost (for example, off-target effects or toxicity on the host) provided by a drug (Fig. 3a). Two general strategies are currently used to combat microbial infections: training the host by vaccination and using small-molecule inhibitors to target the virus or the host. Here we provide perspectives on how common features of noncanonical viral gene expression could serve as a starting point for the development of antiviral therapies.

**Fig. 3: Strategy for therapeutic and prophylactic development of novel antiviral agents.**

---

 **figure3**

**a**, A balance between viral inhibition and host toxicity underlies therapeutic development. Targeting viral-specific functions or host functions that are more important (in a given time frame) to the virus than the host paves the way for the generation of therapeutic agents. **b**, *Cis*-acting nucleic-acid structural elements that are

involved in unconventional viral expression mechanisms (such as pseudoknots in PRFs, and stem loops in polymerase slippage sites and IRES) can be directly targeted by small molecules, host factors and antisense oligonucleotides (AON) or indirectly targeted by modulating the related host factors. **c**, Targeting of virus-specific processes in gene expression (such as cap-snatching and RdRp) that are shared among viruses and not found in hosts offers a high specificity for antiviral agents. The targeting of host dependencies that are used by several viruses provides an alternative route to pan-viral therapeutic agents.

[Full size image](#)

## ARFs as vaccination targets

A goal of vaccination is to generate broadly protective antibodies and/or cross-reactive T cells that are directed against viral targets. However, the design of effective and universal vaccines is often hampered by rapid changes of viral antigens through mutation, recombination or re-assortment. For instance, antigenic drift and shift in the surface glycoproteins of IAV have hampered the development of a universal vaccine against influenza virus<sup>[100](#)</sup>. Thus, a major challenge remains to find ideal vaccination targets that are both highly immunogenic and genetically constrained from mutation owing to potential fitness loss.

ARFs have long been neglected as potential candidates for vaccine or drug development, and might provide a solution to this conundrum. ARFs (such as overprinted ORFs) feature an overall low synonymous divergence<sup>[101,102,103](#)</sup>, and are therefore expected to be relatively constrained from accumulating mutations (as mutations in these regions are likely to disrupt more than one viral protein). Importantly, proteins encoded by ARFs have been shown to be abundantly synthesized during infections<sup>[104,105,106,107](#)</sup> and can be efficiently processed through class-I MHC processing pathways and induce cytotoxic T lymphocyte responses<sup>[108,109,110](#)</sup>.

The use of ARF as epitopes has been proposed for HIV<sup>[108,111,112,113](#)</sup>, influenza virus<sup>[110](#)</sup> and in some cancers<sup>[109](#)</sup> and has several major advantages. First, ARFs in simian immunodeficiency virus and HIV contribute greatly to CD8<sup>+</sup> T cell responses in infected individuals and trigger a stronger cytotoxic T lymphocyte response compared to epitopes that target the canonical proteins<sup>[108,114](#)</sup>. The potential of ARFs as epitopes is further substantiated by the observation that codon-optimized recombinant HIV vaccines (in which ARFs are disrupted or skewed) trigger a reduced cytotoxic T lymphocyte response compared to non-codon optimized vaccines<sup>[112](#)</sup>. Second, cytotoxic T lymphocyte responses to at least some ARF epitopes do not drive viral escape<sup>[113](#)</sup> and presentation of ARF epitopes has been associated with favourable clinical outcomes<sup>[111](#)</sup>. Finally, overprinting ORFs tend to be highly conserved among

strains of the same virus, as in IAV<sup>98</sup>. Taken together, these findings suggest that ARFs and overprinting ORFs present potential antigen candidates for the development of new vaccines and for therapies based on chimeric antigen receptor T cells<sup>115</sup>.

## Targeting viral nucleic acid structures

Many viruses rely on the presence of *cis*-acting structural elements in their genomes for protein expression. These elements tend to be highly conserved, and have both structural and sequence-specific properties; they therefore present excellent targets for drug development (Fig. 3b). These strategies require precise knowledge of the sequence and structure of the nucleic acid target region, as well as its viral and host binding partners.

Structure-targeting drugs can be designed following two strategies. First, a drug can disrupt or alter the structure of a *cis* element. For example, a compound (known as ligand 43) discovered from an in silico small-molecule screen has been shown to specifically inhibit –1 PRFs in SARS-CoV by altering the plasticity of a viral RNA pseudoknot<sup>116,117,118</sup>. Second, a drug can inhibit cofactor binding to a structural element. For example, benzimidazole (a potential inhibitor of hepatitis C virus (HCV)<sup>119,120</sup>) functions by widening the interhelical angle in the viral internal ribosomal entry site (IRES), which results in reduced interaction with ribosome subunits and thus the inhibition of translation<sup>121,122</sup>.

In theory, the high conservation at structure and sequence levels makes viral *cis* elements ideal targets for antisense oligonucleotides, which work by disrupting structure formation or induce degradation of the RNA by recruitment of RNase H. Indeed, the first drug approved by the US Food and Drug Administration (fomivirsen) for treating cytomegalovirus retinitis in individuals infected with HIV is an antisense drug. Several other antisense-based antiviral drugs against HIV, HCV, ebolavirus and Marburg virus have entered clinical trials. However, antisense oligonucleotide technology has some caveats. Besides considerations of delivery method (which have previously been reviewed<sup>123</sup>), virus escape can occur. For example, an antisense oligonucleotide inhibitor (ISI-14803) of HCV that targets the IRES has been shown to exert selective pressure on the IRES sequence<sup>124,125</sup>. This resulted in mutations accumulating in the virus in patients during a phase-I clinical trial, although no mutations were detected at the antisense oligonucleotide binding site<sup>124</sup>. Taken together, these data suggest that the design of drugs based on antisense oligonucleotides

requires a careful analysis of the surrounding structures. Alternatively, it may be necessary to use multiplex delivery of antisense oligonucleotides (that is, to target several regions of the structure at the same time), such that compensatory escape mutations will be unable to take hold.

## Targeting virus-specific mechanisms of gene expression

Many viruses rely on their own proxies of host enzymes (for example, the capping machinery of the *Coronaviridae*) or pathways (for example, the cap-snatching of the *Orthomyxoviridae*) to express viral proteins (Box 2). Inhibitory drugs against these virus-specific proteins and pathways should achieve high specificity for the virus with minimal effect on the host (Fig. 3a).

Cap-snatching, which is used only by influenza viruses and other segmented negative-sense viruses, presents one such targetable pathway. To date, at least three small-molecule antiviral agents (favipiravir, pimodivir and baloxavir) that target the PB1, PB2 and PA subunits, respectively, of the influenza viral polymerase trimer have entered clinical development (as has previously been reviewed<sup>126</sup>). Baloxavir has been approved for treating influenza virus infections in the USA and Japan, and was generated through rational design against the cap-dependent endonuclease active site of the IAV PA protein<sup>127</sup>. Baloxavir has been shown to effectively inhibit cap-snatching activities in both IAV and influenza B virus<sup>127</sup>, and has broader antiviral effects than current standard-of-care anti-influenza drugs<sup>128,129</sup>. Success with these drugs may pave the way for the development of antiviral agents against other highly pathogenic cap-snatching viruses.

Conserved protein domains across viral families might provide targets for broader-acting antiviral agents (Fig. 3c). For example, RdRp is essential to RNA viruses and shares a similar 3D structural conformation<sup>130</sup> and mechanism of action across species, which suggests that drugs that target RdRp could have activities in different viral families. Favipiravir—which was initially discovered on the basis of its antiviral activity against IAV—has been shown to exhibit antiviral activity against other RNA viruses, including viruses that cause fatal haemorrhagic fevers (arenaviruses, peribunyaviruses and filoviruses)<sup>131</sup>.

Although viral-targeting drugs offer high specificity, a potential issue is the acquisition of drug-resistant mutations in the viral targets. In the case of baloxavir, IAV recovered from 1.1 to 19.5% of patients treated with the drug developed up to 138 compensatory mutations<sup>[132](#)</sup>. A possible solution is combination therapy: because the targets of combination therapy are often located in different pathways or proteins, it is more difficult for the viral to acquire resistance compared to monotherapies. Indeed, combination therapies have been shown to slow down the acquisition of resistance and yield effective viral clearance<sup>[133](#)</sup>, as exemplified by the combinatorial ‘highly active antiretroviral therapy’ (HAART) used in controlling HIV infections<sup>[134](#)</sup>, as well as similar strategies using in the treatments of cancers<sup>[135](#)</sup> and multidrug-resistant bacterial infections (as has previously been reviewed<sup>[136](#)</sup>).

Unfortunately most drugs—whether developed by academic or commercial institutions—are developed as single agents, and face a range of legal and regulatory issues that might hamper their use in the testing of combination therapies. Thus, a shift in drug-development paradigms towards a more collaborative environment among research bodies and clinicians is imperative for the future development of combinatorial strategies.

## **Box 2** Viral proxies of host molecules

Many host processes are encoded by large multimodal complexes that are confined to specific subcellular compartments. Viruses use proxies of host machinery and processes that can accordingly be considered as component-level and process-level proxies.

### **Component-level proxies**

Viruses may encode their own simplified versions of host proteins. For instance, cellular RNA capping occurs co-transcriptionally in the nucleus by a series of capping enzymes through removal of a monophosphate from 5' mRNA and the subsequent transferring and methylation of a GMP. The m<sup>7</sup>G cap is a critical modification for viral mRNA, as it protects viral mRNA from degradation and allows translation. Although some viruses can use host capping enzymes, others have evolved to express viral substitutes to

these host enzymes. In most cases, this entails simplified and/or non-canonical proteins (as compared to their host counterparts). The *Mononegavirales* rely on an unconventional multifunctional enzyme known as L protein that serves as both RdRp and capping enzyme. In this process, a covalent mRNA–enzyme intermediate is first formed between the 5' monophosphorylated mRNA and the L protein<sup>[181](#)</sup>. This mRNA–enzyme intermediate is then transferred to a GDP receptor. Subsequently, two methylation modifications sequentially occur on the ribose-2'-O position of the first nucleotide and on the guanine N-7 position of the cap<sup>[181](#)</sup>. Another unconventional capping mechanism is adopted by *Alphaviridae* (as has previously been reviewed<sup>[182](#)</sup>), in which pre-methylated GTP is donated to 5' diphosphorylated mRNA, forming the cap<sup>[183](#)</sup>. Other viruses have innovated structural mimics of the 5' cap. For example, picornaviruses encode the protein Vpg, which covalently links to the 5' of the genome and mimics m<sup>7</sup>G caps. The full spectrum of this strategy has previously been reviewed<sup>[184](#)</sup>.

## Process-level proxies

Viruses may evolve completely unique strategies that are analogous to host processes. For example, RNA splicing typically occurs co-transcriptionally in the nucleus of the host. Viruses (such as coronaviruses and other viruses in the order *Nidovirales*) that replicate in the cytoplasm therefore do not have ready access to the splicing machinery of the host. Instead, these positive-strand viruses use a mechanism of discontinuous transcription to generate a nested set of minus-strand subgenomic mRNAs. These subgenomic mRNAs all share a leader sequence derived from the 5' end of the viral genome and serve as templates for mRNA production. Discontinuous transcription is regulated by transcription regulating sequences. Transcription regulating sequences mediate long-range RNA–RNA interactions that promote viral polymerase template-switching during transcription (as previously reviewed<sup>[185](#)</sup>). Notably, a recent analysis of SARC-CoV-2 indicates that many unidentified ORFs can be generated by discontinuous transcription<sup>[186,187](#)</sup>. Discontinuous transcription can thus be considered as different from alternative splicing mechanistically, but analogous to splicing in terms of the end result (that is, the generation of several transcripts from one gene or genomic region). Similarly, polyprotein

processing is commonly adopted by RNA viruses and retroviruses in which a polyprotein is expressed from a single RNA species and subsequently cleaved by viral or host proteases into functional proteins. These unique viral enzymes or pathways, and the host factors that might regulate these processes, represent potential targets in strategies for viral eradication.

## Host dependencies as targets of pan-viral therapies

Although the high mutation rates of viruses suggest an unlimited evolutionary potential, a virus that is fully co-adapted to its host will have very few neutral sites in its genome<sup>[137](#)</sup>—which locks the virus into evolutionary stasis and limits marked divergence over the long term. In support of this, an analysis of HBV genomes recovered from prehistoric periods has shown that these viruses were only 1.3–3% divergent from modern circulating strains<sup>[138,139](#)</sup>. This suggests that a viable strategy for antiviral development can be achieved by targeting host dependencies, which can result from indirect or direct interactions between a virus and its host (Fig. [3c](#)).

When considering the inhibition of a host dependency a trade-off exists between viral inhibition and the potential disruption of host cellular functions. A parallel can be observed with cancer therapeutic agents: cancer cells that are heavily reliant on essential host functions can be killed by short-term or partial inhibition against these functions (for example, topoisomerase or proteasome inhibitors), while maintaining minimal long-term damage to the patient. The ideal therapeutic targets for viral infections would be host factors upon which viruses heavily depend, and the short-term or partial inhibition of which over the course of an infection is well-tolerated by the host. Furthermore, if commonalities in host dependencies exist among different viruses, targeting these dependencies might allow the development of broad-spectrum or pan-viral therapeutic agents. This could contribute to combating newly emerging infections that lack efficient antiviral therapies (for example, as in the current COVID-19 pandemic).

### Direct dependencies

Viral proteins or RNA may directly interact with host factors to give rise to direct dependencies. The identification of direct host dependencies requires knowledge of host–viral protein–protein and protein–nucleic acid interactions that are shared and important among different viral families. The inhibition of these proteins or processes is therefore likely to have broad-spectrum antiviral effects.

Several viral species require a common set of host factors (collectively known as the IRES *trans*-acting factors) for viral IRES translation. The inhibition of these factors therefore blocks replication of viruses from several unrelated families. For example, the inhibition of the host ribosome-binding protein receptor for activated C kinase 1 (which is co-opted by many viruses in IRES-mediated translation<sup>140</sup>) effectively inhibited HCV and herpes simplex virus infection with no significant effect on the viability or proliferation of the human host cells<sup>140,141</sup>.

Another host dependency is protein localization to the endoplasmic reticulum, which is shared by several evolutionarily distant viruses such IAV, HIV and dengue virus<sup>142</sup>. As predicted, treatment with small-molecule inhibitors of SEC61 (a protein complex that mediates co-translational translocation in endoplasmic reticulum and endoplasmic reticulum–Golgi intermediate compartments) showed suppression of replication of all three of these viruses *in vitro*<sup>142</sup>. Different iterations of SEC61 inhibitors have been shown to effectively suppress Zika virus and coronavirus replication *in vitro*<sup>143,144</sup>. Further work is needed to evaluate their activity *in vivo*, but the underlying general concept is that viruses have a strong requirement—in a small temporal window of active infection—for oxidative folding and modification associated with apical trafficking<sup>142,143,144</sup>. Along similar lines, host glycosylation enzymes (which are extensively used for viral surface protein modification) have inspired the development of vaccines and therapeutic agents—for example, the use of glycans as vaccine adjuvants for HIV<sup>145,146</sup> and antiviral drugs (zanamivir and oseltamivir) for IAV.

## Indirect dependencies

An indirect host dependency arises from indirect functional interactions between the virus and a host protein or process. One example of such a dependency is the importance of the host splicing machinery for viruses that replicate in cytosol. For instance, infections with SARS-CoV-2 have been shown to cause a marked increase in spliceosome components in host cells<sup>147</sup>. Viruses can disrupt host splicing function by triggering nucleocytoplasmic translocation and the sequestering of spliceosome components (in the case of rotavirus<sup>148,149</sup>, which has previously been reviewed<sup>150</sup>) or by inducing changes in splicing patterns of host cellular genes (in the case of influenza virus<sup>149</sup>, Zika virus<sup>151</sup>, human cytomegalovirus<sup>152</sup>, and in hepatitis B virus- and HCV-related hepatocellular carcinoma<sup>153</sup>).

The therapeutic targeting of alternative splicing by small molecules or protein inhibitors and antisense oligonucleotides has been proposed in the treatment of cancer, on the basis of the observation of pro-oncogenic isoforms generated by defective alternative splicing (as previously reviewed<sup>154,155</sup>). Altering the splice pattern of a receptor for viral entry using antisense oligonucleotides could generate a decoy receptor and prevent infection. Overall, the pervasive involvement of host splicing machinery in viral gene expression suggests that modulation of splicing might serve as a promising antiviral therapeutic strategy.

## Conclusion

Viruses use a diverse array of noncanonical transcriptional and translational strategies to greatly expand the coding potential of, and add novel functionality to, their small genomes. However, to do so they have relied on unique enzymatic activities or become dependent on host functions. Viral enzymes that have no homology with human enzymes represent ideal targets for the development of virus-specific inhibitors. Host dependencies are also valuable targets as—in many cases—these dependencies exist broadly across different viruses. We surmise that future developments in our

biochemical and detailed mechanistic understanding of how viruses make proteins will inform the development of therapeutic agents and vaccines.

# References

1. 1.

Kozak, M. Point mutations define a sequence flanking the AUG initiator codon that modulates translation by eukaryotic ribosomes. *Cell* **44**, 283–292 (1986).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

2. 2.

Chung, W. Y., Wadhawan, S., Szklarczyk, R., Pond, S. K. & Nekrutenko, A. A first look at ARFome: dual-coding genes in mammalian genomes. *PLoS Comput. Biol.* **3**, e91 (2007).

[ADS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

3. 3.

Michel, A. M. et al. Observation of dually decoded regions of the human genome using ribosome profiling data. *Genome Res.* **22**, 2219–2229 (2012).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

4. 4.

Schlub, T. E. & Holmes, E. C. Properties and abundance of overlapping genes in viruses. *Virus Evol.* **6**, veaa009 (2020).

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

5. 5.

Biswas, P., Jiang, X., Pacchia, A. L., Dougherty, J. P. & Peltz, S. W. The human immunodeficiency virus type 1 ribosomal frameshifting

site is an invariant sequence determinant and an important target for antiviral therapy. *J. Virol.* **78**, 2082–2087 (2004).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

6. 6.

Shehu-Xhilaga, M., Crowe, S. M. & Mak, J. Maintenance of the Gag/Gag-Pol ratio is important for human immunodeficiency virus type 1 RNA dimerization and viral infectivity. *J. Virol.* **75**, 1834–1841 (2001).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

7. 7.

Jacks, T., Madhani, H. D., Masiarz, F. R. & Varmus, H. E. Signals for ribosomal frameshifting in the Rous sarcoma virus Gag-Pol region. *Cell* **55**, 447–458 (1988). **This article reports the discovery of the frameshift site and stem-loop structure as requirements for PRFs in the synthesis of Gag–Pol protein.**

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

8. 8.

van Knippenberg, I., Carlton-Smith, C. & Elliott, R. M. The N-terminus of Bunyamwera orthobunyavirus NSs protein is essential for interferon antagonism. *J. Gen. Virol.* **91**, 2002–2006 (2010).

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

9. 9.

McFadden, N. et al. Norovirus regulation of the innate immune response and apoptosis occurs via the product of the alternative open reading frame 4. *PLoS Pathog.* **7**, e1002413 (2011).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

10. 10.

Scholthof, H. B. The Tombusvirus-encoded P19: from irrelevance to elegance. *Nat. Rev. Microbiol.* **4**, 405–411 (2006).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

11. 11.

Chen, W. et al. A novel influenza A virus mitochondrial protein that induces cell death. *Nat. Med.* **7**, 1306–1312 (2001).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

12. 12.

Rancurel, C., Khosravi, M., Dunker, A. K., Romero, P. R. & Karlin, D. Overlapping genes produce proteins with unusual sequence properties and offer insight into de novo protein creation. *J. Virol.* **83**, 10719–10736 (2009). **This paper was one of the first to suggest that overprinting is a mechanism for de novo gene and protein creation, and uses viruses as a model system to characterize the properties and the phylogenetic distributions of such proteins.**

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

13. 13.

Neme, R. & Tautz, D. Phylogenetic patterns of emergence of new genes support a model of frequent de novo evolution. *BMC Genomics* **14**, 117 (2013).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

14. 14.

Penno, C., Kumari, R., Baranov, P. V., van Sinderen, D. & Atkins, J. F. Specific reverse transcriptase slippage at the HIV ribosomal frameshift

sequence: potential implications for modulation of GagPol synthesis. *Nucleic Acids Res.* **45**, 10156–10167 (2017).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

15. 15.

Hausmann, S., Garcin, D., Delenda, C. & Kolakofsky, D. The versatility of paramyxovirus RNA polymerase stuttering. *J. Virol.* **73**, 5568–5576 (1999).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

16. 16.

Barr, J. N. & Wertz, G. W. Polymerase slippage at vesicular stomatitis virus gene junctions to generate poly(A) is regulated by the upstream 3'-AUAC-5' tetranucleotide: implications for the mechanism of transcription termination. *J. Virol.* **75**, 6901–6913 (2001).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

17. 17.

Thomas, S. M., Lamb, R. A. & Paterson, R. G. Two mRNAs that differ by two nontemplated nucleotides encode the amino c-terminal proteins P and V of the paramyxovirus SV5. *Cell* **54**, 891–902 (1988).  
**A seminal paper that first identified transcriptional slippage as a mechanism of generating multiple transcripts from one gene in paramyxovirus SV5.**

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

18. 18.

Olspert, A., Chung, B. Y., Atkins, J. F., Carr, J. P. & Firth, A. E. Transcriptional slippage in the positive-sense RNA virus family Potyviridae. *EMBO Rep.* **16**, 995–1004 (2015).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

19. 19.

Ratinier, M. et al. Transcriptional slippage prompts recoding in alternate reading frames in the hepatitis C virus (HCV) core sequence from strain HCV-1. *J. Gen. Virol.* **89**, 1569–1578 (2008).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

20. 20.

Kolakofsky, D., Roux, L., Garcin, D. & Ruigrok, R. W. H. Paramyxovirus mRNA editing, the “rule of six” and error catastrophe: a hypothesis. *J. Gen. Virol.* **86**, 1869–1877 (2005).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

21. 21.

Dillon, P. J. & Parks, G. D. Role for the phosphoprotein P subunit of the paramyxovirus polymerase in limiting induction of host cell antiviral responses. *J. Virol.* **81**, 11116–11127 (2007).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

22. 22.

Gainey, M. D., Dillon, P. J., Clark, K. M., Manuse, M. J. & Parks, G. D. Paramyxovirus-induced shutoff of host and viral protein synthesis: role of the P and V proteins in limiting PKR activation. *J. Virol.* **82**, 828–839 (2008).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

23. 23.

Didcock, L., Young, D. F., Goodbourn, S. & Randall, R. E. The V protein of simian virus 5 inhibits interferon signalling by targeting

STAT1 for proteasome-mediated degradation. *J. Virol.* **73**, 9928–9933 (1999).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

24. 24.

Kato, A., Kiyotani, K., Sakai, Y., Yoshida, T. & Nagai, Y. The paramyxovirus, Sendai virus, V protein encodes a luxury function required for viral pathogenesis. *EMBO J.* **16**, 578–587 (1997).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

25. 25.

Vulliémoz, D. & Roux, L. “Rule of six”: how does the Sendai virus RNA polymerase keep count? *J. Virol.* **75**, 4506–4518 (2001).

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

26. 26.

Iseni, F. et al. Chemical modification of nucleotide bases and mRNA editing depend on hexamer or nucleoprotein phase in Sendai virus nucleocapsids. *RNA* **8**, 1056–1067 (2002).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

27. 27.

Shabman, R. S. et al. Deep sequencing identifies noncanonical editing of Ebola and Marburg virus RNAs in infected cells. *MBio* **5**, e02011-14 (2014).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

28. 28.

Mehedi, M. et al. A new Ebola virus nonstructural glycoprotein expressed through RNA editing. *J. Virol.* **85**, 5406–5414 (2011).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

29. 29.

Sanchez, A., Trappier, S. G., Mahy, B. W., Peters, C. J. & Nichol, S. T. The virion glycoproteins of Ebola viruses are encoded in two reading frames and are expressed through transcriptional editing. *Proc. Natl Acad. Sci. USA* **93**, 3602–3607 (1996).

[ADS](#) [CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

30. 30.

Sanchez, A. et al. Biochemical analysis of the secreted and virion glycoproteins of Ebola virus. *J. Virol.* **72**, 6442–6447 (1998).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

31. 31.

Volchkov, V. E. et al. GP mRNA of Ebola virus is edited by the Ebola virus polymerase and by T7 and vaccinia virus polymerases. *Virology* **214**, 421–430 (1995).

[CAS](#) [Google Scholar](#)

32. 32.

Selman, M., Dankar, S. K., Forbes, N. E., Jia, J. J. & Brown, E. G. Adaptive mutation in influenza A virus non-structural gene is linked to host switching and induces a novel protein by alternative splicing. *Emerg. Microbes Infect.* **1**, e42 (2012).

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

33. 33.

Neumann, G., Hughes, M. T. & Kawaoka, Y. Influenza A virus NS2 protein mediates vRNP nuclear export through NES-independent interaction with hCRM1. *EMBO J.* **19**, 6751–6758 (2000).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

34. 34.

Tsai, P. L. et al. Cellular RNA binding proteins NS1-BP and hnRNP K regulate influenza A virus RNA splicing. *PLoS Pathog.* **9**, e1003460 (2013).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

35. 35.

Shih, S. R. & Krug, R. M. Novel exploitation of a nuclear function by influenza virus: the cellular SF2/ASF splicing factor controls the amount of the essential viral M2 ion channel protein in infected cells. *EMBO J.* **15**, 5415–5427 (1996).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

36. 36.

Artarini, A. et al. Regulation of influenza A virus mRNA splicing by CLK1. *Antiviral Res.* **168**, 187–196 (2019).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

37. 37.

Bogdanow, B. et al. The dynamic proteome of influenza A virus infection identifies M segment splicing as a host range determinant. *Nat. Commun.* **10**, 5518 (2019).

[ADS](#) [CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

38. 38.

Jiang, T., Nogales, A., Baker, S. F., Martinez-Sobrido, L. & Turner, D. H. Mutations designed by ensemble defect to misfold conserved RNA structures of influenza A segments 7 and 8 affect splicing and attenuate viral replication in cell culture. *PLoS ONE* **11**, e0156906 (2016).

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

39. 39.

Cai, Z. et al. VirusCircBase: a database of virus circular RNAs. *Brief Bioinform.* **22**, 2182–2190 (2021).

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

40. 40.

Zhao, J. et al. Transforming activity of an oncoprotein-encoding circular RNA from human papillomavirus. *Nat. Commun.* **10**, 2300 (2019). **This paper reports how oncogenic strains of human papillomavirus use back-splicing to generate circular RNAs of oncogene E7, which have an essential role in the malignant transformation of infected cells.**

[ADS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

41. 41.

Atkins, J. F., Loughran, G., Bhatt, P. R., Firth, A. E. & Baranov, P. V. Ribosomal frameshifting and transcriptional slippage: from genetic steganography and cryptography to adventitious use. *Nucleic Acids Res.* **44**, 7007–7078 (2016).

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

42. 42.

Loughran, G., Firth, A. E. & Atkins, J. F. Ribosomal frameshifting into an overlapping gene in the 2B-encoding region of the cardiovirus

genome. *Proc. Natl Acad. Sci. USA* **108**, E1111–E1119 (2011).

[ADS](#) [CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

43. 43.

Harger, J. W., Meskauskas, A. & Dinman, J. D. An “integrated model” of programmed ribosomal frameshifting. *Trends Biochem. Sci.* **27**, 448–454 (2002).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

44. 44.

Dinman, J. D. Mechanisms and implications of programmed translational frameshifting. *Wiley Interdiscip. Rev. RNA* **3**, 661–673 (2012).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

45. 45.

Jacks, T. et al. Characterization of ribosomal frameshifting in HIV-1 Gag-Pol expression. *Nature* **331**, 280–283 (1988).

[ADS](#) [CAS](#) [Google Scholar](#)

46. 46.

Dulude, D., Baril, M. & Brakier-Gingras, L. Characterization of the frameshift stimulatory signal controlling a programmed –1 ribosomal frameshift in the human immunodeficiency virus type 1. *Nucleic Acids Res.* **30**, 5094–5102 (2002).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

47. 47.

Cassan, M., Delaunay, N., Vaquero, C. & Rousset, J. P. Translational frameshifting at the Gag-Pol junction of human immunodeficiency virus type 1 is not increased in infected T-lymphoid cells. *J. Virol.* **68**, 1501–1508 (1994).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

48. 48.

Brierley, I., Meredith, M. R., Bloys, A. J. & Hagervall, T. G. Expression of a coronavirus ribosomal frameshift signal in *Escherichia coli*: influence of tRNA anticodon modification on frameshifting. *J. Mol. Biol.* **270**, 360–373 (1997).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

49. 49.

Firth, A. E., Blitvich, B. J., Wills, N. M., Miller, C. L. & Atkins, J. F. Evidence for ribosomal frameshifting and a novel overlapping gene in the genomes of insect-specific flaviviruses. *Virology* **399**, 153–166 (2010).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

50. 50.

Baranov, P. V. et al. Programmed ribosomal frameshifting in decoding the SARS-CoV genome. *Virology* **332**, 498–510 (2005).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

51. 51.

Plant, E. P., Rakauskaite, R., Taylor, D. R. & Dinman, J. D. Achieving a golden mean: mechanisms by which coronaviruses ensure synthesis of the correct stoichiometric ratios of viral proteins. *J. Virol.* **84**, 4330–4340 (2010).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

52. 52.

Su, M. C., Chang, C. T., Chu, C. H., Tsai, C. H. & Chang, K. Y. An atypical RNA pseudoknot stimulator and an upstream attenuation signal for  $-1$  ribosomal frameshifting of SARS coronavirus. *Nucleic Acids Res.* **33**, 4265–4275 (2005).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

53. 53.

Dulude, D., Berchiche, Y. A., Gendron, K., Brakier-Gingras, L. & Heveker, N. Decreasing the frameshift efficiency translates into an equivalent reduction of the replication of the human immunodeficiency virus type 1. *Virology* **345**, 127–136 (2006).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

54. 54.

Garcia-Miranda, P. et al. Stability of HIV frameshift site RNA correlates with frameshift efficiency and decreased virus infectivity. *J. Virol.* **90**, 6906–6917 (2016).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

55. 55.

Karacostas, V., Wolffe, E. J., Nagashima, K., Gonda, M. A. & Moss, B. Overexpression of the HIV-1 Gag-Pol polyprotein results in intracellular activation of HIV-1 protease and inhibition of assembly and budding of virus-like particles. *Virology* **193**, 661–671 (1993).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

56. 56.

Suzuki, Y. et al. Characterization of RyDEN (C19orf66) as an interferon-stimulated cellular inhibitor against dengue virus replication. *PLoS Pathog.* **12**, e1005357 (2016).

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

57. 57.

Wang, X. et al. Regulation of HIV-1 Gag-Pol expression by shiftless, an inhibitor of programmed –1 ribosomal frameshifting. *Cell* **176**, 625–635.e14 (2019). **This article describes how the host factor shiftless functions as a universal inhibitor of –1 PRFs by causing premature translational termination in HIV-1 and several other viruses.**

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

58. 58.

Schwartz, S., Felber, B. K., Fenyö, E. M. & Pavlakis, G. N. Env and Vpu proteins of human immunodeficiency virus type 1 are produced from multiple bicistronic mRNAs. *J. Virol.* **64**, 5448–5456 (1990). **The authors report the discovery of leaky scanning of an upstream viral protein as the mechanism of generating the envelope protein in HIV-1.**

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

59. 59.

Kolakofsky, D., Le Mercier, P., Iseni, F. & Garcin, D. Viral DNA polymerase scanning and the gymnastics of Sendai virus RNA synthesis. *Virology* **318**, 463–473 (2004).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

60. 60.

Stacey, S. N. et al. Leaky scanning is the predominant mechanism for translation of human papillomavirus type 16 E7 oncoprotein from E6/E7 bicistronic mRNA. *J. Virol.* **74**, 7284–7297 (2000).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

61. 61.

Fuller, F., Bhowm, A. S. & Bishop, D. H. Bunyavirus nucleoprotein, N, and a non-structural protein, NSS, are coded by overlapping reading frames in the S RNA. *J. Gen. Virol.* **64**, 1705–1714 (1983).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

62. 62.

Vera-Otarola, J. et al. The Andes hantavirus NSs protein is expressed from the viral small mRNA by a leaky scanning mechanism. *J. Virol.* **86**, 2176–2187 (2012).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

63. 63.

Strelbel, K., Klimkait, T. & Martin, M. A. A novel gene of HIV-1, *vpu*, and its 16-kilodalton product. *Science* **241**, 1221–1223 (1988).

[ADS](#) [CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

64. 64.

Willey, R. L., Maldarelli, F., Martin, M. A. & Strelbel, K. Human immunodeficiency virus type 1 Vpu protein induces rapid degradation of CD4. *J. Virol.* **66**, 7193–7200 (1992).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

65. 65.

Terwilliger, E. F., Cohen, E. A., Lu, Y. C., Sodroski, J. G. & Haseltine, W. A. Functional role of human immunodeficiency virus type 1 Vpu. *Proc. Natl Acad. Sci. USA* **86**, 5163–5167 (1989).

[ADS](#) [CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

66. 66.

Sauter, D. et al. Tetherin-driven adaptation of Vpu and Nef function and the evolution of pandemic and nonpandemic HIV-1 strains. *Cell Host Microbe* **6**, 409–421 (2009).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

67. 67.

Wise, H. M. et al. Overlapping signals for translational regulation and packaging of influenza A virus segment 2. *Nucleic Acids Res.* **39**, 7775–7790 (2011).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

68. 68.

Zell, R. et al. Prevalence of PB1-F2 of influenza A viruses. *J. Gen. Virol.* **88**, 536–546 (2007).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

69. 69.

McAuley, J. L. et al. PB1-F2 proteins from H5N1 and 20<sup>th</sup> century pandemic influenza viruses cause immunopathology. *PLoS Pathog.* **6**, e1001014 (2010).

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

70. 70.

Zamarin, D., García-Sastre, A., Xiao, X., Wang, R. & Palese, P. Influenza virus PB1-F2 protein induces cell death through mitochondrial ANT3 and VDAC1. *PLoS Pathog.* **1**, e4 (2005).

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

71. 71.

Lulla, V. et al. An upstream protein-coding region in enteroviruses modulates virus infection in gut epithelial cells. *Nat. Microbiol.* **4**, 280–292 (2019).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

72. 72.

Irigoyen, N. et al. High-resolution analysis of coronavirus gene expression by RNA sequencing and ribosome profiling. *PLoS Pathog.* **12**, e1005473 (2016).

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

73. 73.

Wu, H. Y., Guan, B. J., Su, Y. P., Fan, Y. H. & Brian, D. A. Reselection of a genomic upstream open reading frame in mouse hepatitis coronavirus 5'-untranslated-region mutants. *J. Virol.* **88**, 846–858 (2014).

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

74. 74.

Hofmann, M. A., Senanayake, S. D. & Brian, D. A. A translation-attenuating intraleader open reading frame is selected on coronavirus mRNAs during persistent infection. *Proc. Natl Acad. Sci. USA* **90**, 11733–11737 (1993).

[ADS](#) [CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

75. 75.

Shabman, R. S. et al. An upstream open reading frame modulates ebola virus polymerase translation and virus replication. *PLoS Pathog.* **9**, e1003147 (2013). **This paper relates how an upstream ORF suppresses the translation of the downstream canonical L protein as a way to maintain protein expression level in ebolavirus.**

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

76. 76.

Degnin, C. R., Schleiss, M. R., Cao, J. & Geballe, A. P. Translational inhibition mediated by a short upstream open reading frame in the human cytomegalovirus gpUL4 (gp48) transcript. *J. Virol.* **67**, 5514–5521 (1993).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

77. 77.

Chen, A., Kao, Y. F. & Brown, C. M. Translation of the first upstream ORF in the hepatitis B virus pregenomic RNA modulates translation at the core and polymerase initiation codons. *Nucleic Acids Res.* **33**, 1169–1181 (2005).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

78. 78.

Calvo, S. E., Pagliarini, D. J. & Mootha, V. K. Upstream open reading frames cause widespread reduction of protein expression and are polymorphic among humans. *Proc. Natl Acad. Sci. USA* **106**, 7507–7512 (2009).

[ADS](#) [CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

79. 79.

Ji, Z., Song, R., Regev, A. & Struhl, K. Many lncRNAs, 5'UTRs, and pseudogenes are translated and some are likely to express functional proteins. *eLife* **4**, e08890 (2015).

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

80. 80.

Jin, X., Turcott, E., Englehardt, S., Mize, G. J. & Morris, D. R. The two upstream open reading frames of oncogene *mdm2* have different translational regulatory properties. *J. Biol. Chem.* **278**, 25716–25721 (2003).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

81. 81.

Firth, A. E. & Brierley, I. Non-canonical translation in RNA viruses. *J. Gen. Virol.* **93**, 1385–1409 (2012).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

82. 82.

Young, S. K. & Wek, R. C. Upstream open reading frames differentially regulate gene-specific translation in the integrated stress response. *J. Biol. Chem.* **291**, 16927–16935 (2016).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

83. 83.

Gupta, K. C. & Patwardhan, S. ACG, the initiator codon for a Sendai virus protein. *J. Biol. Chem.* **263**, 8553–8556 (1988). **This paper represents one of the first observations of near-cognate start codon use in viral protein expression.**

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

84. 84.

Boeck, R., Curran, J., Matsuoka, Y., Compans, R. & Kolakofsky, D. The parainfluenza virus type 1 P/C gene uses a very efficient GUG codon to start its C' protein. *J. Virol.* **66**, 1765–1768 (1992).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

85. 85.

Prats, A. C., De Billy, G., Wang, P. & Darlix, J. L. CUG initiation codon used for the synthesis of a cell surface antigen coded by the murine leukemia virus. *J. Mol. Biol.* **205**, 363–372 (1989).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

86. 86.

Corcelette, S., Massé, T. & Madjar, J. J. Initiation of translation by non-AUG codons in human T-cell lymphotropic virus type I mRNA encoding both Rex and Tax regulatory proteins. *Nucleic Acids Res.* **28**, 1625–1634 (2000).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

87. 87.

Machkovech, H. M., Bloom, J. D. & Subramaniam, A. R. Comprehensive profiling of translation initiation in influenza virus infected cells. *PLoS Pathog.* **15**, e1007518 (2019).

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

88. 88.

Shirako, Y. Non-AUG translation initiation in a plant RNA virus: a forty-amino-acid extension is added to the N terminus of the soil-borne wheat mosaic virus capsid protein. *J. Virol.* **72**, 1677–1682 (1998).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

89. 89.

Carroll, R. & Derse, D. Translation of equine infectious anemia virus bicistronic *tat-rev* mRNA requires leaky ribosome scanning of the *tat* CTG initiation codon. *J. Virol.* **67**, 1433–1440 (1993).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

90. 90.

Sasaki, J. & Nakashima, N. Methionine-independent initiation of translation in the capsid protein of an insect RNA virus. *Proc. Natl Acad. Sci. USA* **97**, 1512–1515 (2000). **This paper reports the discovery that translation initiation of an insect virus mRNA is independent of both AUG and the methionine initiator tRNA.**

[ADS](#) [CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

91. 91.

Wilson, J. E., Powell, M. J., Hoover, S. E. & Sarnow, P. Naturally occurring dicistronic cricket paralysis virus RNA is regulated by two internal ribosome entry sites. *Mol. Cell. Biol.* **20**, 4990–4999 (2000).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

92. 92.

Rialdi, A. et al. The RNA exosome syncs IAV-RNAPII transcription to promote viral ribogenesis and infectivity. *Cell* **169**, 679–692.e14 (2017).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

93. 93.

Dias, A. et al. The cap-snatching endonuclease of influenza virus polymerase resides in the PA subunit. *Nature* **458**, 914–918 (2009).

[ADS](#) [CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

94. 94.

1. Reich, S. et al. Structural insight into cap-snatching and RNA synthesis by influenza polymerase. *Nature* **516**, 361–366 (2014).

[ADS](#) [CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

2. 95.

Koppstein, D., Ashour, J. & Bartel, D. P. Sequencing the cap-snatching repertoire of H1N1 influenza provides insight into the mechanism of viral transcription initiation. *Nucleic Acids Res.* **43**, 5052–5064 (2015).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

3. 96.

Gu, W. et al. Influenza A virus preferentially snatches noncoding RNA caps. *RNA* **21**, 2067–2075 (2015).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

4. 97.

Sikora, D., Rocheleau, L., Brown, E. G. & Pelchat, M. Influenza A virus cap-snatches host RNAs based on their abundance early after infection. *Virology* **509**, 167–177 (2017).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

5. 98.

Ho, J. S. Y. et al. Hybrid gene origination creates human-virus chimeric proteins during infection. *Cell* **181**, 1502–1517 (2020). **This article describes the discovery of upstream AUGs in cap-snatched host sequences being transcribed and translated to generate host-virus hybrid proteins.**

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

6. 99.

Wise, H. M. et al. An alternative AUG codon that produces an N-terminally extended form of the influenza A virus NP is a virulence factor for a swine-derived virus. Preprint at <https://doi.org/10.1101/738427> (2019).

7. 100.

Kim, H., Webster, R. G. & Webby, R. J. Influenza virus: dealing with a drifting and shifting pathogen. *Viral Immunol.* **31**, 174–183 (2018).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

8. 101.

Firth, A. E. & Brown, C. M. Detecting overlapping coding sequences in virus genomes. *BMC Bioinformatics* **7**, 75 (2006).

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

9. 102.

Firth, A. E. & Brown, C. M. Detecting overlapping coding sequences with pairwise alignments. *Bioinformatics* **21**, 282–292 (2005).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

10. 103.

Jagger, B. W. et al. An overlapping protein-coding region in influenza A virus segment 3 modulates the host response. *Science* **337**, 199–204 (2012).

[ADS](#) [CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

11. 104.

Mohamadi, M. et al. Hepatitis C virus alternative reading frame protein (ARFP): production, features, and pathogenesis. *J. Med. Virol.* **92**, 2930–2937 (2020).

12. 105.

Zanker, D. J. et al. Influenza A virus infection induces viral and cellular defective ribosomal products encoded by alternative reading frames. *J. Immunol.* **202**, 3370–3380 (2019).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

13. 106.

Depledge, D. P. et al. Direct RNA sequencing on nanopore arrays redefines the transcriptional complexity of a viral pathogen. *Nat. Commun.* **10**, 754 (2019).

[ADS](#) [CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

14. 107.

Di, H. et al. Expanded subgenomic mRNA transcriptome and coding capacity of a nidovirus. *Proc. Natl Acad. Sci. USA* **114**, E8895–E8904 (2017).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

15. 108.

Bansal, A. et al. CD8 T cell response and evolutionary pressure to HIV-1 cryptic epitopes derived from antisense transcription. *J. Exp. Med.* **207**, 51–59 (2010). **This paper describes how an ARF in HIV encodes cryptic epitopes that contribute to the majority of the CD8 T cell response during infection.**

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

16. 109.

Vetsika, E. K. et al. Sequential administration of the native TERT<sub>572</sub> cryptic peptide enhances the immune response initiated by its optimized variant TERT<sub>572Y</sub> in cancer patients. *J. Immunother.* **34**, 641–650 (2011).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

17. 110.

Dolan, B. P., Li, L., Takeda, K., Bennink, J. R. & Yewdell, J. W. Defective ribosomal products are the major source of antigenic peptides endogenously generated from influenza A virus neuraminidase. *J. Immunol.* **184**, 1419–1424 (2010).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

18. 111.

Bansal, A. et al. Enhanced recognition of HIV-1 cryptic epitopes restricted by HLA class I alleles associated with a favorable clinical outcome. *J. Acquir. Immune Defic. Syndr.* **70**, 1–8 (2015).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

19. 112.

Bet, A., Sterrett, S., Sato, A., Bansal, A. & Goepfert, P. A. Characterization of T-cell responses to cryptic epitopes in recipients of a noncodon-optimized HIV-1 vaccine. *J. Acquir. Immune Defic. Syndr.* **65**, 142–150 (2014).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

20. 113.

Peng, B. J. et al. Antisense-derived HIV-1 cryptic epitopes are not major drivers of viral evolution during the acute phase of infection. *J. Virol.* **92**, e00711-18 (2018).

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

21. 114.

Maness, N. J. et al. Robust, vaccine-induced CD8<sup>+</sup> T lymphocyte response against an out-of-frame epitope. *J. Immunol.* **184**, 67–72 (2010). **Four articles<sup>111–114</sup> demonstrate how the T cell response to cryptic epitopes produced from ARFs in HIV is robust, associated with improved clinical outcomes and associated with minimal viral escape.**

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

22. 115.

Seif, M., Einsele, H. & Löffler, J. CAR T cells beyond cancer: hope for immunomodulatory therapy of infectious diseases. *Front. Immunol.* **10**, 2711 (2019).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

23. 116.

Park, S. J., Kim, Y. G. & Park, H. J. Identification of RNA pseudoknot-binding ligand that inhibits the -1 ribosomal frameshifting of SARS-coronavirus by structure-based virtual screening. *J. Am. Chem. Soc.* **133**, 10094–10100 (2011). **This paper describes how an RNA structure-based inhibitor screening revealed a candidate that targets and reduces the efficiency of –1 PRFs in SARS-CoV.**

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

24. 117.

Ritchie, D. B., Soong, J., Sikkema, W. K. & Woodside, M. T. Anti-frameshifting ligand reduces the conformational plasticity of the SARS virus pseudoknot. *J. Am. Chem. Soc.* **136**, 2196–2199 (2014).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

25. 118.

Ritchie, D. B., Foster, D. A. & Woodside, M. T. Programmed –1 frameshifting efficiency correlates with RNA pseudoknot conformational plasticity, not resistance to mechanical unfolding. *Proc. Natl Acad. Sci. USA* **109**, 16167–16172 (2012).

[ADS](#) [CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

26. 119.

Seth, P. P. et al. SAR by MS: discovery of a new class of RNA-binding small molecules for the hepatitis C virus: internal ribosome entry site IIA subdomain. *J. Med. Chem.* **48**, 7099–7102 (2005).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

27. 120.

Parsons, J. et al. Conformational inhibition of the hepatitis C virus internal ribosome entry site RNA. *Nat. Chem. Biol.* **5**, 823–825 (2009).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

28. 121.

Dibrov, S. M. et al. Structure of a hepatitis C virus RNA domain in complex with a translation inhibitor reveals a binding mode reminiscent of riboswitches. *Proc. Natl Acad. Sci. USA* **109**, 5223–5228 (2012).

[ADS](#) [CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

29. 122.

Yamamoto, H. et al. Molecular architecture of the ribosome-bound hepatitis C virus internal ribosomal entry site RNA. *EMBO J.* **34**,

3042–3058 (2015).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

30. 123.

Juliano, R., Alam, M. R., Dixit, V. & Kang, H. Mechanisms and strategies for effective delivery of antisense and siRNA oligonucleotides. *Nucleic Acids Res.* **36**, 4158–4171 (2008).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

31. 124.

Soler, M., McHutchison, J. G., Kwoh, T. J., Dorr, F. A. & Pawlotsky, J. M. Virological effects of ISIS 14803, an antisense oligonucleotide inhibitor of hepatitis C virus (HCV) internal ribosome entry site (IRES), on HCV IRES in chronic hepatitis C patients and examination of the potential role of primary and secondary HCV resistance in the outcome of treatment. *Antivir. Ther.* **9**, 953–968 (2004).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

32. 125.

Zhang, H. et al. Antisense oligonucleotide inhibition of hepatitis C virus (HCV) gene expression in livers of mice infected with an HCV–vaccinia virus recombinant. *Antimicrob. Agents Chemother.* **43**, 347–353 (1999).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

33. 126.

Hayden, F. G. & Shindo, N. Influenza virus polymerase inhibitors in clinical development. *Curr. Opin. Infect. Dis.* **32**, 176–186 (2019).

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

34. 127.

Noshi, T. et al. In vitro characterization of baloxavir acid, a first-in-class cap-dependent endonuclease inhibitor of the influenza virus polymerase PA subunit. *Antiviral Res.* **160**, 109–117 (2018).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

35. 128.

Fukao, K. et al. Combination treatment with the cap-dependent endonuclease inhibitor baloxavir marboxil and a neuraminidase inhibitor in a mouse model of influenza A virus infection. *J. Antimicrob. Chemother.* **74**, 654–662 (2019).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

36. 129.

Mishin, V. P. et al. Susceptibility of influenza A, B, C, and D viruses to baloxavir. *Emerg. Infect. Dis.* **25**, 1969–1972 (2019).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

37. 130.

de Farias, S. T., Dos Santos Junior, A. P., Rêgo, T. G. & José, M. V. Origin and evolution of RNA-dependent RNA polymerase. *Front. Genet.* **8**, 125 (2017).

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

38. 131.

Guedj, J. et al. Antiviral efficacy of favipiravir against ebola virus: a translational study in cynomolgus macaques. *PLoS Med.* **15**, e1002535 (2018).

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

39. 132.

Omoto, S. et al. Characterization of influenza virus variants induced by treatment with the endonuclease inhibitor baloxavir marboxil. *Sci. Rep.* **8**, 9633 (2018).

[ADS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

40. 133.

Hofmann, W. P., Soriano, V. & Zeuzem, S. in *Antiviral Strategies: Handbook of Experimental Pharmacology* (eds Kräusslich H. G. & Bartenschlager, R.) 321–346 (2009).

41. 134.

Arts, E. J. & Hazuda, D. J. HIV-1 antiretroviral drug therapy. *Cold Spring Harb. Perspect. Med.* **2**, a007161 (2012).

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

42. 135.

Bayat Mokhtari, R. et al. Combination therapy in combating cancer. *Oncotarget* **8**, 38022–38043 (2017).

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

43. 136.

Worthington, R. J. & Melander, C. Combination approaches to combat multidrug-resistant bacteria. *Trends Biotechnol.* **31**, 177–184 (2013).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

44. 137.

Simmonds, P., Aiewsakun, P. & Katzourakis, A. Prisoners of war – host adaptation and its constraints on virus evolution. *Nat. Rev.*

*Microbiol.* **17**, 321–328 (2019).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

45. 138.

Krause-Kyora, B. et al. Neolithic and medieval virus genomes reveal complex evolution of hepatitis B. *eLife* **7**, e36666 (2018).

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

46. 139.

Mühlemann, B. et al. Ancient hepatitis B viruses from the Bronze Age to the Medieval period. *Nature* **557**, 418–423 (2018). **Two articles<sup>138,139</sup> show that hepatitis B virus isolated from samples from the Bronze and Neolithic Age is minimally divergent from modern strains, which suggests that viral adaptation to the host may confer an upper limit to the mutations that can be accumulated over time.**

[ADS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

47. 140.

Majzoub, K. et al. RACK1 controls IRES-mediated translation of viruses. *Cell* **159**, 1086–1095 (2014).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

48. 141.

Ullah, H., Hou, W., Dakshanamurthy, S. & Tang, Q. Host targeted antiviral (HTA): functional inhibitor compounds of scaffold protein RACK1 inhibit herpes simplex virus proliferation. *Oncotarget* **10**, 3209–3226 (2019).

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

49. 142.

Heaton, N. S. et al. Targeting viral proteostasis limits influenza virus, HIV, and dengue virus infection. *Immunity* **44**, 46–58 (2016). **This paper reports that viruses exhibit host dependency of SEC61-mediated protein translocation and folding, and that inhibition of SEC61 provides a broad-spectrum inhibition on growth and infectivity in several viruses.**

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

50. 143.

Shah, P. S. et al. Comparative flavivirus–host protein interaction mapping reveals mechanisms of dengue and Zika virus pathogenesis. *Cell* **175**, 1931–1945 (2018).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

51. 144.

Gordon, D. E. et al. A SARS-CoV-2 protein interaction map reveals targets for drug repurposing. *Nature* **583**, 459–468 (2020).

[ADS](#) [CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

52. 145.

Dalziel, M., Crispin, M., Scanlan, C. N., Zitzmann, N. & Dwek, R. A. Emerging principles for the therapeutic exploitation of glycosylation. *Science* **343**, 1235681 (2014).

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

53. 146.

Watanabe, Y., Bowden, T. A., Wilson, I. A. & Crispin, M. Exploitation of glycosylation in enveloped virus pathobiology. *Biochim. Biophys. Acta, Gen. Subj.* **1863**, 1480–1497 (2019).

[CAS](#) [Google Scholar](#)

54. 147.

Bojkova, D. et al. Proteomics of SARS-CoV-2-infected host cells reveals therapy targets. *Nature* **583**, 469–472 (2020).

[ADS](#) [CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

55. 148.

Dhillon, P. et al. Cytoplasmic relocalization and colocalization with viroplasms of host cell proteins, and their role in rotavirus infection. *J. Virol.* **92**, e00612-18 (2018).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

56. 149.

Zhao, N. et al. Influenza virus infection causes global RNAPII termination defects. *Nat. Struct. Mol. Biol.* **25**, 885–893 (2018).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

57. 150.

Boudreault, S., Roy, P., Lemay, G. & Bisaillon, M. Viral modulation of cellular RNA alternative splicing: a new key player in virus-host interactions? *Wiley Interdiscip. Rev. RNA* **10**, e1543 (2019).

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

58. 151.

Bonenfant, G. et al. Zika virus subverts stress granules to promote and restrict viral gene expression. *J. Virol.* **93**, e00520-19 (2019).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

59. 152.

Batra, R. et al. RNA-binding protein CPEB1 remodels host and viral RNA landscapes. *Nat. Struct. Mol. Biol.* **23**, 1101–1110 (2016).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

60. 153.

Tremblay, M. P. et al. Global profiling of alternative RNA splicing events provides insights into molecular differences between various types of hepatocellular carcinoma. *BMC Genomics* **17**, 683 (2016).

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

61. 154.

Lin, J. C. Therapeutic applications of targeted alternative splicing to cancer treatment. *Int. J. Mol. Sci.* **19**, E75 (2017).

[Google Scholar](#)

62. 155.

Urbanski, L. M., Leclair, N. & Anczuków, O. Alternative-splicing defects in cancer: splicing regulators and their downstream targets, guiding the way to novel cancer therapeutics. *Wiley Interdiscip. Rev. RNA* **9**, e1476 (2018).

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

63. 156.

Pelletier, J. & Sonenberg, N. Internal initiation of translation of eukaryotic mRNA directed by a sequence derived from poliovirus RNA. *Nature* **334**, 320–325 (1988).

[ADS](#) [CAS](#) [Google Scholar](#)

64. 157.

Jang, S. K. et al. A segment of the 5' nontranslated region of encephalomyocarditis virus RNA directs internal entry of ribosomes during in vitro translation. *J. Virol.* **62**, 2636–2643 (1988). **This paper represents one of the first observations of IRESs as an alternative cap-independent translation mechanism that is used by encephalomyocarditis virus.**

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

65. 158.

Griffiths, A. & Coen, D. M. An unusual internal ribosome entry site in the herpes simplex virus thymidine kinase gene. *Proc. Natl Acad. Sci. USA* **102**, 9667–9672 (2005).

[ADS](#) [CAS](#) [Google Scholar](#)

66. 159.

Kang, S. T., Wang, H. C., Yang, Y. T., Kou, G. H. & Lo, C. F. The DNA virus white spot syndrome virus uses an internal ribosome entry site for translation of the highly expressed nonstructural protein ICP35. *J. Virol.* **87**, 13263–13278 (2013).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

67. 160.

Zhao, J. et al. IRESbase: a comprehensive database of experimentally validated internal ribosome entry sites. *Genomics Proteomics Bioinformatics* **18**, 129–139 (2020).

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

68. 161.

Sweeney, T. R., Abaeva, I. S., Pestova, T. V. & Hellen, C. U. The mechanism of translation initiation on type 1 picornavirus IRESSs. *EMBO J.* **33**, 76–92 (2014).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

69. 162.

Kolupaeva, V. G., Lomakin, I. B., Pestova, T. V. & Hellen, C. U. Eukaryotic initiation factors 4G and 4A mediate conformational changes downstream of the initiation codon of the encephalomyocarditis virus internal ribosomal entry site. *Mol. Cell. Biol.* **23**, 687–698 (2003).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

70. 163.

Fernández, I. S., Bai, X. C., Murshudov, G., Scheres, S. H. & Ramakrishnan, V. Initiation of translation by cricket paralysis virus IRES requires its translocation in the ribosome. *Cell* **157**, 823–831 (2014).

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

71. 164.

Deniz, N., Lenarcic, E. M., Landry, D. M. & Thompson, S. R. Translation initiation factors are not required for Dicistroviridae IRES function in vivo. *RNA* **15**, 932–946 (2009).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

72. 165.

Fütterer, J., Kiss-László, Z. & Hohn, T. Nonlinear ribosome migration on cauliflower mosaic virus 35S RNA. *Cell* **73**, 789–802 (1993).

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

73. 166.

Pooggin, M. M., Fütterer, J., Skryabin, K. G. & Hohn, T. A short open reading frame terminating in front of a stable hairpin is the conserved feature in pregenomic RNA leaders of plant pararetroviruses. *J. Gen. Virol.* **80**, 2217–2228 (1999).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

74. 167.

Yueh, A. & Schneider, R. J. Selective translation initiation by ribosome jumping in adenovirus-infected and heat-shocked cells. *Genes Dev.* **10**, 1557–1567 (1996).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

75. 168.

Latorre, P., Kolakofsky, D. & Curran, J. Sendai virus Y proteins are initiated by a ribosomal shunt. *Mol. Cell. Biol.* **18**, 5021–5031 (1998).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

76. 169.

Meyers, G. Translation of the minor capsid protein of a calicivirus is initiated by a novel termination-dependent reinitiation mechanism. *J. Biol. Chem.* **278**, 34051–34060 (2003).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

77. 170.

Horvath, C. M., Williams, M. A. & Lamb, R. A. Eukaryotic coupled translation of tandem cistrons: identification of the influenza B virus BM2 polypeptide. *EMBO J.* **9**, 2639–2647 (1990).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

78. 171.

Ahmadian, G., Randhawa, J. S. & Easton, A. J. Expression of the ORF-2 protein of the human respiratory syncytial virus M2 gene is initiated by a ribosomal termination-dependent reinitiation mechanism. *EMBO J.* **19**, 2681–2689 (2000).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

79. 172.

Gould, P. S. & Easton, A. J. Coupled translation of the respiratory syncytial virus M2 open reading frames requires upstream sequences. *J. Biol. Chem.* **280**, 21972–21980 (2005).

[CAS](#) [Google Scholar](#)

80. 173.

Jeudy, S., Abergel, C., Claverie, J. M. & Legendre, M. Translation in giant viruses: a unique mixture of bacterial and eukaryotic termination schemes. *PLoS Genet.* **8**, e1003122 (2012).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

81. 174.

Schueren, F. & Thoms, S. Functional translational readthrough: a systems biology perspective. *PLoS Genet.* **12**, e1006196 (2016).

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

82. 175.

Leinfelder, W., Zehelein, E., Mandrand-Berthelot, M. A. & Böck, A. Gene for a novel tRNA species that accepts l-serine and cotranslationally inserts selenocysteine. *Nature* **331**, 723–725 (1988).

[ADS](#) [CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

83. 176.

Lee, B. J., Worland, P. J., Davis, J. N., Stadtman, T. C. & Hatfield, D. L. Identification of a selenocysteyl-tRNA(Ser) in mammalian cells that recognizes the nonsense codon, UGA. *J. Biol. Chem.* **264**, 9724–9727 (1989).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

84. 177.

Ryan, M. D. & Drew, J. Foot-and-mouth disease virus 2A oligopeptide mediated cleavage of an artificial polyprotein. *EMBO J.* **13**, 928–933 (1994).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

85. 178.

Donnelly, M. L. L. et al. Analysis of the aphthovirus 2A/2B polyprotein ‘cleavage’ mechanism indicates not a proteolytic reaction, but a novel translational effect: a putative ribosomal ‘skip’. *J. Gen. Virol.* **82**, 1013–1025 (2001).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

86. 179.

Luke, G. A. et al. Occurrence, function and evolutionary origins of ‘2A-like’ sequences in virus genomes. *J. Gen. Virol.* **89**, 1036–1042 (2008).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

87. 180.

Sharma, P. et al. 2A peptides provide distinct solutions to driving stop-carry on translational recoding. *Nucleic Acids Res.* **40**, 3143–3151 (2012).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

88. 181.

Ogino, T. & Banerjee, A. K. Unconventional mechanism of mRNA capping by the RNA-dependent RNA polymerase of vesicular stomatitis virus. *Mol. Cell* **25**, 85–97 (2007).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

89. 182.

Decroly, E., Ferron, F., Lescar, J. & Canard, B. Conventional and unconventional mechanisms for capping viral mRNA. *Nat. Rev. Microbiol.* **10**, 51–65 (2011).

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

90. 183.

Ahola, T. & Kääriäinen, L. Reaction in alphavirus mRNA capping: formation of a covalent complex of nonstructural protein nsP1 with 7-methyl-GMP. *Proc. Natl Acad. Sci. USA* **92**, 507–511 (1995).

[ADS](#) [CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

91. 184.

Goodfellow, I. The genome-linked protein VPg of vertebrate viruses – a multifaceted protein. *Curr. Opin. Virol.* **1**, 355–362 (2011).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

92. 185.

Sola, I., Almazán, F., Zúñiga, S. & Enjuanes, L. Continuous and discontinuous RNA synthesis in coronaviruses. *Annu. Rev. Virol.* **2**, 265–288 (2015).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

93. 186.

Kim, D. et al. The architecture of SARS-CoV-2 transcriptome. *Cell* **181**, 914–921.e10 (2020).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

94. 187.

Finkel, Y. et al. The coding capacity of SARS-CoV-2. *Nature* **589**, 125–130 (2021).

[ADS](#) [CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

[Download references](#)

## **Acknowledgements**

I.M. is supported by Burroughs Wellcome Fund 1017892, NIH/NIAID-U01AI150748 and Chan Zuckerberg Initiative 2018-191895

## **Author information**

### Author notes

1. These authors contributed equally: Jessica Sook Yuin Ho, Zeyu Zhu

## Affiliations

1. Department of Microbiology, Icahn School of Medicine at Mount Sinai, New York, NY, USA

Jessica Sook Yuin Ho, Zeyu Zhu & Ivan Marazzi

2. Global Health and Emerging Pathogens Institute, Icahn School of Medicine at Mount Sinai, New York, NY, USA

Ivan Marazzi

## Authors

1. Jessica Sook Yuin Ho

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

2. Zeyu Zhu

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

3. Ivan Marazzi

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

## Contributions

J.S.Y.H., Z.Z. and I.M. wrote the manuscript.

## Corresponding author

Correspondence to [Ivan Marazzi](#).

# Ethics declarations

## Competing interests

The authors declare no competing interests.

## Additional information

**Peer review information** *Nature* thanks Sean Whelan and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

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

## Rights and permissions

[Reprints and Permissions](#)

## About this article



Check for  
updates

## Cite this article

Ho, J.S.Y., Zhu, Z. & Marazzi, I. Unconventional viral gene expression mechanisms as therapeutic targets. *Nature* **593**, 362–371 (2021).  
<https://doi.org/10.1038/s41586-021-03511-5>

[Download citation](#)

- Received: 08 June 2020

- Accepted: 22 March 2021
- Published: 19 May 2021
- Issue Date: 20 May 2021
- DOI: <https://doi.org/10.1038/s41586-021-03511-5>

## Comments

By submitting a comment you agree to abide by our [Terms](#) and [Community Guidelines](#). If you find something abusive or that does not comply with our terms or guidelines please flag it as inappropriate.

[Download PDF](#)

Advertisement

---

This article was downloaded by **calibre** from <https://www.nature.com/articles/s41586-021-03511-5>

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

- Perspective
- [Published: 19 May 2021](#)

# The data-driven future of high-energy-density physics

- [Peter W. Hatfield](#) [ORCID: orcid.org/0000-0002-3065-457X<sup>1</sup>](#),
- [Jim A. Gaffney](#) [ORCID: orcid.org/0000-0002-2408-0047<sup>2</sup>](#),
- [Gemma J. Anderson](#) [ORCID: orcid.org/0000-0002-4169-5553<sup>2</sup>](#),
- [Suzanne Ali](#) [ORCID: orcid.org/0000-0003-1823-3788<sup>2</sup>](#),
- [Luca Antonelli](#) [ORCID: orcid.org/0000-0003-0694-948X<sup>3</sup>](#),
- [Suzan Bağışmez du Pree](#) [ORCID: orcid.org/0000-0003-1599-6974<sup>4</sup>](#),
- [Jonathan Citrin<sup>5</sup>](#),
- [Marta Fajardo<sup>6</sup>](#),
- [Patrick Knapp<sup>7</sup>](#),
- [Brendan Kettle](#) [ORCID: orcid.org/0000-0003-4424-4448<sup>8</sup>](#),
- [Bogdan Kustowski](#) [ORCID: orcid.org/0000-0003-2085-3206<sup>2</sup>](#),
- [Michael J. MacDonald](#) [ORCID: orcid.org/0000-0002-6295-6978<sup>2</sup>](#),
- [Derek Mariscal<sup>2</sup>](#),
- [Madison E. Martin](#) [ORCID: orcid.org/0000-0003-0218-8206<sup>2</sup>](#),
- [Taisuke Nagayama<sup>7</sup>](#),
- [Charlotte A. J. Palmer](#) [ORCID: orcid.org/0000-0002-3762-0380<sup>9</sup>](#),
- [J. Luc Peterson](#) [ORCID: orcid.org/0000-0002-5167-5708<sup>2</sup>](#),
- [Steven Rose](#) [ORCID: orcid.org/0000-0001-6808-6355<sup>1,8</sup>](#),
- [J J Ruby<sup>10</sup>](#),
- [Carl Shneider](#) [ORCID: orcid.org/0000-0002-3689-6959<sup>11</sup>](#),
- [Matt J. V. Streeter](#) [ORCID: orcid.org/0000-0001-9086-9831<sup>8</sup>](#),
- [Will Trickey](#) [ORCID: orcid.org/0000-0001-7013-4395<sup>3</sup>](#) &
- [Ben Williams<sup>12</sup>](#)

- 857 Accesses
- 43 Altmetric
- [Metrics details](#)

## Subjects

- [Laboratory astrophysics](#)
- [Laser-produced plasmas](#)
- [Nuclear fusion and fission](#)
- [Scientific data](#)

## Abstract

High-energy-density physics is the field of physics concerned with studying matter at extremely high temperatures and densities. Such conditions produce highly nonlinear plasmas, in which several phenomena that can normally be treated independently of one another become strongly coupled. The study of these plasmas is important for our understanding of astrophysics, nuclear fusion and fundamental physics—however, the nonlinearities and strong couplings present in these extreme physical systems makes them very difficult to understand theoretically or to optimize experimentally. Here we argue that machine learning models and data-driven methods are in the process of reshaping our exploration of these extreme systems that have hitherto proved far too nonlinear for human researchers. From a fundamental perspective, our understanding can be improved by the way in which machine learning models can rapidly discover complex interactions in large datasets. From a practical point of view, the newest generation of extreme physics facilities can perform experiments multiple times a second (as opposed to approximately daily), thus moving away from human-based control towards automatic control based on real-time interpretation of diagnostic data and updates of the physics model. To make the most of these emerging opportunities, we suggest proposals for the community in terms of research design, training, best practice and support for synthetic diagnostics and data analysis.

## Access options

Subscribe to Journal

Get full journal access for 1 year

\$199.00

only \$3.90 per issue

[Subscribe](#)

All prices are NET prices.

VAT will be added later in the checkout.

Tax calculation will be finalised during checkout.

Rent or Buy article

Get time limited or full article access on ReadCube.

from \$8.99

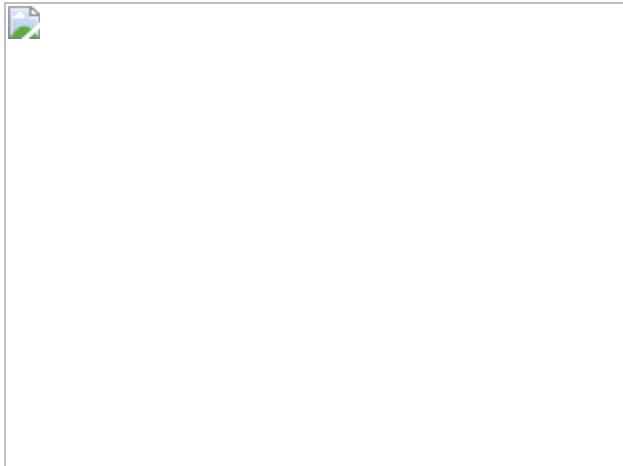
[Rent or Buy](#)

All prices are NET prices.

### Additional access options:

- [Log in](#)
- [Access through your institution](#)
- [Learn about institutional subscriptions](#)

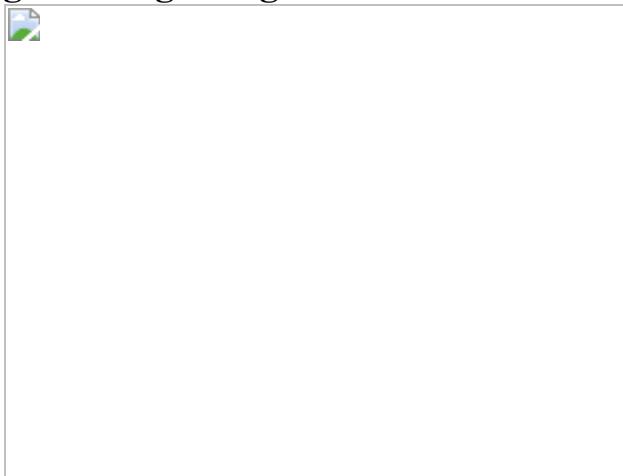
**Fig. 1: Shot rates and energy of large high-powered laser facilities in different eras.**



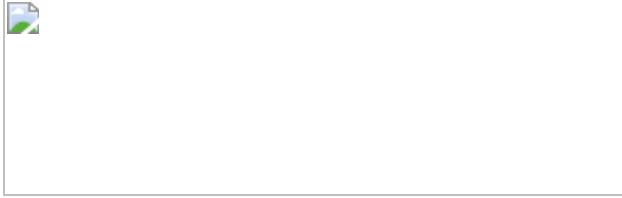
**Fig. 2: Integration of astrophysical information.**



**Fig. 3: Integrating information sources in ICF studies.**



**Fig. 4: High-repetition workflow.**



## References

1. 1.

Rose, S. Set the controls for the heart of the Sun. *Contemp. Phys.* **45**, 109–121 (2004).

[ADS](#) [CAS](#) [Google Scholar](#)

2. 2.

Spears, B. K. et al. Deep learning: a guide for practitioners in the physical sciences. *Phys. Plasmas* **25**, 080901 (2018). **This tutorial paper gives an introduction to scientific machine learning, with examples taken from ICF research.**

[ADS](#) [Google Scholar](#)

3. 3.

Wang, Z., Peterson, J. L., Rea, C. & Humphreys, D. Special issue on machine learning, data science, and artificial intelligence in plasma research. *IEEE Trans. Plasma Sci.* **48**, 1–2 (2020).

[ADS](#) [Google Scholar](#)

4. 4.

Colvin, J. & Larsen, J. *Extreme Physics* (Cambridge Univ. Press, 2013).

5. 5.

Graziani, F., Desjarlais, M. P., Redmer, R. & Trickey, S. B. (eds) *Frontiers and Challenges in Warm Dense Matter* Lecture Notes in Computational Science and Engineering Vol. 96 (Springer, 2014).

6. 6.

Gould, O., Mangles, S., Rajantie, A., Rose, S. & Xie, C. Observing thermal Schwinger pair production. *Phys. Rev. A* **99**, 052120 (2019).

[ADS](#) [CAS](#) [Google Scholar](#)

7. 7.

Millot, M. et al. Experimental evidence for superionic water ice using shock compression. *Nat. Phys.* **14**, 297–302 (2018).

[CAS](#) [Google Scholar](#)

8. 8.

Celliers, P. M. et al. Insulator–metal transition in dense fluid deuterium. *Science* **361**, 677–682 (2018).

[ADS](#) [CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

9. 9.

Joshi, C. & Malka, V. Focus on laser- and beam-driven plasma accelerators. *New J. Phys.* **12**, 045003 (2010).

[ADS](#) [Google Scholar](#)

10. 10.

Hidding, B., Foster, B., Hogan, M. J., Muggli, P. & Rosenzweig, J. B. Directions in plasma wakefield acceleration. *Phil. Trans. R. Soc. Math. Phys. Eng. Sci.* **377**, 20190215 (2019).

[ADS](#) [CAS](#) [Google Scholar](#)

11. 11.

Wang, W.-M. et al. Collimated ultrabright gamma rays from electron wiggling along a petawatt laser-irradiated wire in the QED regime. *Proc. Natl Acad. Sci. USA* **115**, 9911–9916 (2018).

[ADS](#) [CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

12. 12.

Badziak, J. Laser-driven ion acceleration: methods, challenges and prospects. *J. Phys. Conf. Ser.* **959**, 012001 (2018).

[Google Scholar](#)

13. 13.

Feng, J. et al. High-efficiency neutron source generation from photonuclear reactions driven by laser plasma accelerator. *High Energy Density Phys.* **36**, 100753 (2020).

[CAS](#) [Google Scholar](#)

14. 14.

Hidding, B. et al. Plasma wakefield accelerator research 2019–2040: a community-driven UK roadmap compiled by the plasma wakefield accelerator steering committee (PWASC). Working Paper (2019); preprint at <https://arxiv.org/abs/1904.09205>.

15. 15.

Nuckolls, J., Wood, L., Thiessen, A. & Zimmerman, G. Laser compression of matter to super-high densities: thermonuclear applications. *Nature* **239**, 139–142 (1972).

[ADS](#) [CAS](#) [Google Scholar](#)

16. 16.

Hurricane, O. A. & Program, I. Overview of progress and future prospects in indirect drive implosions on the National Ignition Facility. *J. Phys. Conf. Ser.* **717**, 012005 (2016).

[Google Scholar](#)

17. 17.

Fournier, K. *NIF Monthly Highlights for July 2019*. LLNL Technical Report, LLNL-TR-785259; NIF-1006466629981592, <http://www.osti.gov/servlets/purl/1548376/> (Lawrence Livermore National Laboratory, 2019).

18. 18.

Remington, B. A., Rudd, R. E. & Wark, J. S. From microjoules to megajoules and kilobars to gigabars: probing matter at extreme states of deformation. *Phys. Plasmas* **22**, 090501 (2015).

[ADS](#) [Google Scholar](#)

19. 19.

Trines, R. *CLF Annual Report 2017–2018*, <https://www.clf.stfc.ac.uk/Pages/Annual-Report-2017-18.aspx> (Science and Technology Facilities Council, 2019).

20. 20.

Sturm, C. & Stöcker, H. The facility for antiproton and ion research FAIR. *Phys. Part. Nucl. Lett.* **8**, 865–868 (2011).

[Google Scholar](#)

21. 21.

MacDonald, M. J. et al. Measurement of high-dynamic range X-ray Thomson scattering spectra for the characterization of nano-plasmas at LCLS. *Rev. Sci. Instrum.* **87**, 11E709 (2016).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

22. 22.

Mitchell, T. *Machine Learning* (McGraw-Hill, 1997).

23. 23.

Sivia, D. & Skilling, J. *Data Analysis: A Bayesian Tutorial* (Oxford Univ. Press).

24. 24.

Brunton, S. L. & Kutz, J. N. *Data-Driven Science and Engineering* (Cambridge Univ. Press, 2019).

25. 25.

Reichstein, M. et al. Deep learning and process understanding for data-driven Earth system science. *Nature* **566**, 195–204 (2019).

[ADS](#) [CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

26. 26.

Fleming, S. W. & Gupta, H. V. The physics of river prediction. *Phys. Today* **73**, 46–52 (2020).

[Google Scholar](#)

27. 27.

Streeter, M. J. V. et al. Temporal feedback control of high-intensity laser pulses to optimize ultrafast heating of atomic clusters. *Appl. Phys. Lett.* **112**, 244101 (2018).

[ADS](#) [Google Scholar](#)

28. 28.

GopalaSwamy, V. et al. Tripled yield in direct-drive laser fusion through statistical modelling. *Nature* **565**, 581–586 (2019). **This paper is the first to use data-driven approaches to motivate and carry out new ICF experiments.**

[ADS](#) [CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

29. 29.

Hatfield, P. W., Rose, S. J. & Scott, R. H. H. The blind implosion-maker: automated inertial confinement fusion experiment design. *Phys. Plasmas* **26**, 062706 (2019).

[ADS](#) [Google Scholar](#)

30. 30.

Martin, M., London, R., Goluoglu, S. & Whitley, H. An automated design process for short pulse laser driven opacity experiments. *High Energy Density Phys.* **26**, 26–37 (2018).

[ADS](#) [CAS](#) [Google Scholar](#)

31. 31.

Raghu, M. & Schmidt, E. A Survey of Deep Learning for Scientific Discovery (2020); preprint at <http://arxiv.org/abs/2003.11755>.

32. 32.

Thayer, J. et al. Data systems for the Linac Coherent Light Source. *Adv. Struct. Chem. Imag.* **3**, 3 (2017).

[CAS](#) [Google Scholar](#)

33. 33.

Bernal, J. L., Peacock, J. A., Bernal, J. L. & Peacock, J. A. Conservative cosmology: combining data with allowance for unknown

systematics. *J. Cosmol. Astroparticle Phys.* **2018**, 002 (2018). **This paper considers how to get realistic uncertainty estimates in the presence of unknown systematics.**

[Google Scholar](#)

34. 34.

Osthus, D., Vander Wiel, S. A., Hoffman, N. M. & Wysocki, F. J. Prediction uncertainties beyond the range of experience: a case study in inertial confinement fusion implosion experiments. *SIAM/ASA J. Uncertainty Quant.* **7**, 604–633 (2019).

[MathSciNet](#) [MATH](#) [Google Scholar](#)

35. 35.

Lang, M. & Owens, M. J. A variational approach to data assimilation in the solar wind. *Space Weather* **17**, 59–83 (2019).

[ADS](#) [Google Scholar](#)

36. 36.

Gaffney, J. A. et al. Making inertial confinement fusion models more predictive. *Phys. Plasmas* **26**, 082704 (2019).

[ADS](#) [Google Scholar](#)

37. 37.

Kasim, M. F., Galligan, T. P., Topp-Mugglestone, J., Gregori, G. & Vinko, S. M. Inverse problem instabilities in large-scale modeling of matter in extreme conditions. *Phys. Plasmas* **26**, 112706 (2019).

[ADS](#) [Google Scholar](#)

38. 38.

van de Plassche, K. L. et al. Fast modeling of turbulent transport in fusion plasmas using neural networks. *Phys. Plasmas* **27**, 022310 (2020).

[ADS](#) [Google Scholar](#)

39. 39.

Meneghini, O. et al. Self-consistent core-pedestal transport simulations with neural network accelerated models. *Nucl. Fusion* **57**, 086034 (2017).

[ADS](#) [Google Scholar](#)

40. 40.

Anirudh, R., Thiagarajan, J. J., Bremer, P.-T. & Spears, B. K. Improved surrogates in inertial confinement fusion with manifold and cycle consistencies. *Proc. Natl Acad. Sci. USA* **117**, 9741–9746 (2020).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

41. 41.

Kluth, G. et al. Deep learning for NLTE spectral opacities. *Phys. Plasmas* **27**, 052707 (2020). **This paper uses machine learning to emulate opacity calculations, in order to speed up ICF simulations.**

[ADS](#) [CAS](#) [Google Scholar](#)

42. 42.

Humbird, K. D., Peterson, J. L. & McClaren, R. G. Deep neural network initialization with decision trees. *IEEE Trans. Neural Netw. Learn. Syst.* **30**, 1286–1295 (2019). **This paper presents a data-driven method for selecting deep neural network architecture and hyperparameters**

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

43. 43.

Kasim, M. F. et al. Building high accuracy emulators for scientific simulations with deep neural architecture search. Preprint at <https://arxiv.org/abs/2001.08055> (2020).

44. 44.

Kustowski, B. et al. Transfer learning as a tool for reducing simulation bias: application to inertial confinement fusion. *IEEE Trans. Plasma Sci.* **48**, 46–53 (2020). **Transfer learning updates a subset of a pre-trained deep neural network using experimental data; this paper explores its application to sparse ICF datasets.**

[ADS](#) [CAS](#) [Google Scholar](#)

45. 45.

Bambi, C. Astrophysical black holes: a compact pedagogical review. *Ann. Phys.* **530**, 1700430 (2018).

[MathSciNet](#) [Google Scholar](#)

46. 46.

*The National Space Weather Program Strategic Plan.* Technical Report CM-P30–1995, <https://www.icams-portal.gov/publications/spacewx/nswp2.html> (ICAMS, 1995).

47. 47.

Eastwood, J. P. et al. The economic impact of space weather: where do we stand? *Risk Anal.* **37**, 206–218 (2017).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

48. 48.

Camporeale, E. The challenge of machine learning in space weather: nowcasting and forecasting. *Space Weather* **17**, 1166–1207 (2019).

[ADS](#) [Google Scholar](#)

49. 49.

Camporeale, E., Carè, A. & Borovsky, J. E. Classification of solar wind with machine learning. *J. Geophys. Res. Space Phys.* **122**, 10910–10920 (2017).

[ADS](#) [Google Scholar](#)

50. 50.

Chen, Y. et al. Identifying solar flare precursors using time series of SDO/HMI images and SHARP parameters. *Space Weather* **17**, 1404–1426 (2019).

[ADS](#) [Google Scholar](#)

51. 51.

Campi, C. et al. Feature ranking of active region source properties in solar flare forecasting and the uncompromised stochasticity of flare occurrence. *Astrophys. J.* **883**, 150 (2019).

[ADS](#) [Google Scholar](#)

52. 52.

Inceoglu, F. et al. Using machine learning methods to forecast if solar flares will be associated with CMEs and SEPs. *Astrophys. J.* **861**, 128 (2018).

[ADS](#) [Google Scholar](#)

53. 53.

Bobra, M. G. & Ilonidis, S. Predicting coronal mass ejections using machine learning methods. *Astrophys. J.* **821**, 127 (2016).

[ADS](#) [Google Scholar](#)

54. 54.

Sarma, R. et al. Bayesian inference of quasi-linear radial diffusion parameters using Van Allen probes. *J. Geophys. Res. Space Phys.* **125**, e2019JA027618 (2020).

[ADS](#) [Google Scholar](#)

55. 55.

Camporeale, E. et al. A gray-box model for a probabilistic estimate of regional ground magnetic perturbations: enhancing the NOAA operational geospace model with machine learning. *J. Geophys. Res. Space Phys.* **125**, e27684 (2020).

[ADS](#) [Google Scholar](#)

56. 56.

Lamb, K. et al. Correlation of auroral dynamics and GNSS scintillation with an autoencoder. In *Second Workshop on Machine Learning and the Physical Sciences (NeurIPS 2019)* (2019); preprint at <https://arxiv.org/abs/1910.03085>.

57. 57.

Rowlinson, A. et al. Identifying transient and variable sources in radio images. *Astron. Comput.* **27**, 111–129 (2019).

[ADS](#) [Google Scholar](#)

58. 58.

Coughlin, M. W., Dietrich, T., Margalit, B. & Metzger, B. D. Multi-messenger Bayesian parameter inference of a binary neutron-star merger. *Mon. Not. R. Astron. Soc. Lett.* **489**, L91–L96 (2019). **This paper uses both gravitational-wave and multi-wavelength electromagnetic data to simultaneously constrain stellar masses, orbital parameters, and supra-nuclear density equations of state.**

[ADS](#) [Google Scholar](#)

59. 59.

Dorn, C. et al. A generalized Bayesian inference method for constraining the interiors of super Earths and sub-Neptunes. *Astron. Astrophys.* **597**, A37 (2017).

[Google Scholar](#)

60. 60.

Bellinger, E. P. et al. Fundamental parameters of main-sequence stars in an instant with machine learning. *Astrophys. J.* **830**, 31 (2016).

[ADS](#) [Google Scholar](#)

61. 61.

Lochner, M., McEwen, J. D., Peiris, H. V., Lahav, O. & Winter, M. K. Photometric supernova classification with machine learning. *Astrophys. J. Suppl.* **225**, 31 (2016).

[ADS](#) [Google Scholar](#)

62. 62.

Kong, X. et al. Spectral feature extraction for DB white dwarfs through machine learning applied to new discoveries in the SDSS DR12 and DR14. *Publ. Astron. Soc. Pacif.* **130**, 084203 (2018).

[ADS](#) [Google Scholar](#)

63. 63.

Huppenkothen, D., Heil, L. M., Hogg, D. W. & Mueller, A. Using machine learning to explore the long-term evolution of GRS 1915+105. *Mon. Not. R. Astron. Soc.* **466**, 2364–2377 (2017).

[ADS](#) [Google Scholar](#)

64. 64.

Chardin, J. et al. A deep learning model to emulate simulations of cosmic reionization. *Mon. Not. R. Astron. Soc.* **490**, 1055–1065 (2019).

[ADS](#) [CAS](#) [Google Scholar](#)

65. 65.

Saumon, D. & Guillot, T. Shock compression of deuterium and the interiors of Jupiter and Saturn. *Astrophys. J.* **609**, 1170–1180 (2004).

[ADS](#) [CAS](#) [Google Scholar](#)

66. 66.

Smith, R. F. et al. Ramp compression of diamond to five terapascals. *Nature* **511**, 330–333 (2014).

[ADS](#) [CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

67. 67.

Tzeferacos, P. et al. Laboratory evidence of dynamo amplification of magnetic fields in a turbulent plasma. *Nat. Commun.* **9**, 591 (2018).

[ADS](#) [CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

68. 68.

Falcon, R. E. et al. Laboratory measurements of white dwarf photospheric spectral lines: H $\beta$ . *Astrophys. J.* **806**, 214 (2015).

[ADS](#) [Google Scholar](#)

69. 69.

Bailey, J. E. et al. A higher-than-predicted measurement of iron opacity at solar interior temperatures. *Nature* **517**, 56–59 (2015).

[ADS](#) [CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

70. 70.

Nagayama, T. et al. Systematic study of L-shell opacity at stellar interior temperatures. *Phys. Rev. Lett.* **122**, 235001 (2019).

[ADS](#) [CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

71. 71.

de Souza, R. S., Boston, S. R., Coc, A. & Iliadis, C. Thermonuclear fusion rates for tritium + deuterium using Bayesian methods. *Phys. Rev. C* **99**, 014619 (2019).

[ADS](#) [Google Scholar](#)

72. 72.

Zhelavskaya, I. S., Shprits, Y. Y. & Spasojević, M. Empirical modeling of the plasmasphere dynamics using neural networks. *J. Geophys. Res. Space Phys.* **122**, 11227–11244 (2017).

[ADS](#) [Google Scholar](#)

73. 73.

Zhelavskaya, I. S., Shprits, Y. Y. & Spasojevic, M. in *Machine Learning Techniques for Space Weather* 301–327 (Elsevier, 2018).

74. 74.

Freidberg, J. *Plasma Physics and Fusion Energy* (Cambridge Univ. Press, 2008).

75. 75.

Atzeni, S., Meyer-ter Vehn, J. & Meyer-ter Vehn, J. *The Physics of Inertial Fusion: BeamPlasma Interaction, Hydrodynamics, Hot Dense Matter* International Series of Monographs on Physics (Clarendon Press, 2004).

76. 76.

Lindl, J. *Inertial Confinement Fusion: The Quest for Ignition and Energy Gain Using Indirect Drive* (AIP Press, 1998).

77. 77.

Campbell, E. M. & Hogan, W. J. The National Ignition Facility—applications for inertial fusion energy and high-energy-density science. *Plasma Phys. Contr. Fusion* **41**, B39–B56 (1999).

[CAS](#) [Google Scholar](#)

78. 78.

Moses, E. I. Ignition on the National Ignition Facility. *J. Phys. Conf. Ser.* **112**, 012003 (2008).

[Google Scholar](#)

79. 79.

Boehly, T. et al. Initial performance results of the OMEGA laser system. *Opt. Commun.* **133**, 495–506 (1997).

[ADS](#) [CAS](#) [Google Scholar](#)

80. 80.

Slutz, S. A. High-gain magnetized inertial fusion. *Phys. Rev. Lett.* **108**, 025003 (2012).

[ADS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

81. 81.

Gomez, M. R. et al. Experimental demonstration of fusion-relevant conditions in magnetized liner inertial fusion. *Phys. Rev. Lett.* **113**, 155003 (2014).

[ADS](#) [CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

82. 82.

Dimonte, G. Quantitative metrics for evaluating thermonuclear design codes and physics models applied to the National Ignition Campaign. *Phys. Plasmas* **27**, 052709 (2020).

[ADS](#) [CAS](#) [Google Scholar](#)

83. 83.

Yang, C. et al. *Preparing Dense Net for Automated HYDRA Mesh Management via Reinforcement Learning*. Technical Report LLNL-TR-799958, <https://www.osti.gov/servlets/purl/1580017> (Lawrence Livermore National Laboratory, 2019). **This report details the use of deep neural networks to automatically perform adaptive mesh refinement in radiation-hydrodynamics simulations.**

84. 84.

Peterson, J. L. et al. Merlin: enabling machine learning-ready HPC ensembles. Preprint at <https://arxiv.org/abs/1912.02892> (2019).

85. 85.

Peterson, J. L. et al. Zonal flow generation in inertial confinement fusion implosions. *Phys. Plasmas* **24**, 032702 (2017). **This paper presents the first novel ICF design reached using machine learning.**

[ADS](#) [Google Scholar](#)

86. 86.

Amorin, C., Kegelmeyer, L. M. & Kegelmeyer, W. P. A hybrid deep learning architecture for classification of microscopic damage on National Ignition Facility laser optics. *Stat. Analysis Data Mining* **12**, 505–513 (2019). **Here, deep neural network-based image classification is used to detect and classify damage in the NIF laser system, allowing more reliable operation at high energy.**

[MathSciNet](#) [MATH](#) [Google Scholar](#)

87. 87.

Nora, R., Peterson, J. L., Spears, B. K., Field, J. E. & Brandon, S. Ensemble simulations of inertial confinement fusion implosions. *Stat. Analysis Data Mining* **10**, 230–237 (2017).

[MathSciNet](#) [MATH](#) [Google Scholar](#)

88. 88.

Gaffney, J. A. et al. *The JAG Inertial Confinement Fusion Simulation Dataset For Multi-Modal Scientific Deep Learning* <https://library.ucsd.edu/dc/object/bb5534097t> (2020). **This dataset is an example of open data practices in ICF, and is one of the first multi-modal scientific datasets to be released by the HEDP community.**

89. 89.

Thiagarajan, J. J. et al. Designing accurate emulators for scientific processes using calibration-driven deep models. *Nat. Commun.* **11**,

5622 (2020).

[ADS](#) [CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

90. 90.

Hatfield, P. et al. Using sparse Gaussian processes for predicting robust inertial confinement fusion implosion yields. *IEEE Trans. Plasma Sci.* **48**, 14–21 (2020).

[ADS](#) [Google Scholar](#)

91. 91.

Glinsky, M. E. et al. Quantification of MagLIF morphology using the Mallat scattering transformation. *Phys. Plasmas* **27**, 112703 (2019).  
**This paper uses a novel, physics-motivated deep neural network architecture to featurize images of ICF implosions driven on the Z pulsed-power machine.**

[ADS](#) [Google Scholar](#)

92. 92.

Palaniyappan, S. et al. Hydro-scaling of direct-drive cylindrical implosions at the OMEGA and the National Ignition Facility. *Phys. Plasmas* **27**, 042708 (2020).

[ADS](#) [CAS](#) [Google Scholar](#)

93. 93.

Gaffney, J., Clark, D., Sonnad, V. & Libby, S. Development of a Bayesian method for the analysis of inertial confinement fusion experiments on the NIF. *Nucl. Fusion* **53**, 073032 (2013).

[ADS](#) [CAS](#) [Google Scholar](#)

94. 94.

Gaffney, J., Clark, D., Sonnad, V. & Libby, S. Bayesian inference of inaccuracies in radiation transport physics from inertial confinement fusion experiments. *High Energy Density Phys.* **9**, 457–461 (2013).

[ADS](#) [CAS](#) [Google Scholar](#)

95. 95.

Knapp, P. *A Bayesian Parameter Estimation Framework for Understanding Fusion Experiments on Z*. Technical Report SAND2018-1698PE663774, <https://www.osti.gov/biblio/1525597> (2018).

96. 96.

Hsu, A., Cheng, B. & Bradley, P. A. Analysis of NIF scaling using physics informed machine learning. *Phys. Plasmas* **27**, 012703 (2020).

[ADS](#) [CAS](#) [Google Scholar](#)

97. 97.

Humbird, K. D., Peterson, J. L., Spears, B. K. & McClaren, R. G. Transfer learning to model inertial confinement fusion experiments. *IEEE Trans. Plasma Sci.* **48**, 61–70 (2020).

[ADS](#) [Google Scholar](#)

98. 98.

Götzfried, J. et al. Research towards high-repetition rate laser-driven X-ray sources for imaging applications. *Nucl. Instrum. Meth. Phys. Res. A* **909**, 286–289 (2018).

[ADS](#) [Google Scholar](#)

99. 99.

He, Z.-H. et al. Coherent control of plasma dynamics. *Nat. Commun.* **6**, 7156 (2015).

[ADS](#) [CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

100. 100.

Dann, S. J. et al. Laser wakefield acceleration with active feedback at 5 Hz. *Phys. Rev. Accel. Beams* **22**, 041303 (2019).

[ADS](#) [CAS](#) [Google Scholar](#)

101. 101.

Kirschner, J., Mutny, M. M., Hiller, N., Ischebeck, R. & Krause, A. Adaptive and safe Bayesian optimization in high dimensions via one-dimensional subspaces. In *Proc. Machine Learning Res.* **97**, 3429–3438 (2019). **This paper describes a Bayesian optimization method for a free electron laser that prevents the laser from violating any safety constraints.**

[Google Scholar](#)

102. 102.

Shaloo, R. J. et al. Automation and control of laser wakefield accelerators using Bayesian optimization. *Nat. Commun.* **11**, 6355 (2020).

[ADS](#) [CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

103. 103.

Maier, A. R. et al. Decoding sources of energy variability in a laser-plasma accelerator. *Phys. Rev. X* **10**, 031039 (2020).

[CAS](#) [Google Scholar](#)

104. 104.

Team, J. T. et al. Hybrid neural network for density limit disruption prediction and avoidance on J-TEXT tokamak. *Nucl. Fusion* **58**, 056016 (2018).

[ADS](#) [Google Scholar](#)

105. 105.

Fu, Y. et al. Machine learning control for disruption and tearing mode avoidance. *Phys. Plasmas* **27**, 022501 (2020).

[ADS](#) [CAS](#) [Google Scholar](#)

106. 106.

Kates-Harbeck, J., Svyatkovskiy, A. & Tang, W. Predicting disruptive instabilities in controlled fusion plasmas through deep learning. *Nature* **568**, 526–531 (2019).

[ADS](#) [CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

107. 107.

Wilkinson, M. D. et al. Comment: The FAIR guiding principles for scientific data management and stewardship. *Sci. Data* **3**, 160018 (2016).

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

108. 108.

Albertsson, K. et al. Machine learning in high energy physics community white paper. *J. Phys. Conf. Ser.* **1085**, 022008 (2018).

[Google Scholar](#)

109. 109.

Borne, K. D. et al. The revolution in astronomy education: data science for the masses. Preprint at <https://arxiv.org/abs/0909.3895> (2009).

110. 110.

Pasian, F. et al. Science ground segment for the ESA Euclid Mission. In *Software and Cyberinfrastructure for Astronomy II* (eds Radziwill, N. M. & Chiozzi, G.) Vol. 8451, 845104 (SPIE, 2012).

111. 111.

Lyons, L. Discovering the significance of 5 sigma. Preprint at <https://arxiv.org/abs/1310.1284> (2013). **This paper discusses the background, rationale, and advantages and disadvantages of the  $5\sigma$  criterion commonly used in particle physics.**

112. 112.

Roodman, A. Blind analysis in particle physics. In *Proc. Conf. on Statistical Problems in Particle Physics, Astrophysics and Cosmology* (SLAC, 2003).

113. 113.

Roso, L. High repetition rate petawatt lasers. *EPJ Web Conf.* **167**, 01001 (2018).

[Google Scholar](#)

114. 114.

Zheng, W. et al. Laser performance of the SG-III laser facility. In *High Power Laser Science and Engineering* Vol. 4, e21 (Cambridge Univ. Press, 2016).

115. 115.

Danson, C. N. et al. Petawatt and exawatt class lasers worldwide. In *High Power Laser Science and Engineering* Vol. 7, e54 (Cambridge

Univ. Press, 2019). This paper reviews some of the many high-powered lasers in use around the world.

116. 116.

*Opportunities in Intense Ultrafast Lasers*  
<https://doi.org/10.17226/24939> (National Academies Press, 2018).

117. 117.

Lin, Z. et al. SG-II laser elementary research and precision SG-II program. *Fusion Eng. Des.* **44**, 61–66 (1999).

[CAS](#) [Google Scholar](#)

118. 118.

*LULI2000 User Guide*. <https://gargantua.polytechnique.fr/siatel-web/app/linkto/mICYYYTJle5S> (Laboratoire pour l’Utilisation des Lasers Intenses, 2019).

119. 119.

Kirillov, G. A., Murugov, V. M., Punin, V. T. & Shemyakin, V. I. High power laser system ISKRA V. *Laser Part. Beams* **8**, 827–831 (1990).

[ADS](#) [CAS](#) [Google Scholar](#)

120. 120.

Zhao, Z. & Wang, D. XFEL Projects in China. In *Proc. LINAC2018 (Beijing)* <http://accelconf.web.cern.ch/linac2018/html/author.htm> (2019).

121. 121.

Zhang, Z. et al. The laser beamline in SULF facility. In *High Power Laser Science and Engineering* Vol. 8, e4 (Cambridge Univ. Press, 2020).

122. 122.

Schramm, U. et al. First results with the novel petawatt laser acceleration facility in Dresden. In *J. Phys. Conf. Ser.* **874**, 12028 (2017).

[Google Scholar](#)

123. 123.

European XFEL. *Facts and Figures*  
[https://www.xfel.eu/facility/overview/facts\\_and\\_figures/index\\_eng.html](https://www.xfel.eu/facility/overview/facts_and_figures/index_eng.html).

124. 124.

Linac Coherent Light Source. *LCLS-II Design & Performance*  
<https://lcls.slac.stanford.edu/lcls-ii/design-and-performance>.

125. 125.

Yabashi, M., Tanaka, H., Tono, K. & Ishikawa, T. Status of the SACLAC facility. *Appl. Sci.* **7**, 604 (2017).

[Google Scholar](#)

126. 126.

CERN. *FAQ – LHC the Guide*. Technical Report CERN-Brochure-2017-002-Eng, <https://cds.cern.ch/record/2255762> (CERN, 2017).

[Download references](#)

## Acknowledgements

This Perspective is the result of a meeting at the Lorentz Center, University of Leiden, 13–17 January 2020. The Lorentz Centre is funded by the Dutch Research Council (NWO) and the University of Leiden. The meeting also

had support from the John Fell Oxford University Press (OUP) Research Fund. The organizers are grateful to T. Uitbeijerse (Lorentz Center) for facilitating the meeting. P.W.H. acknowledges funding from the Engineering and Physical Sciences Research Council. A portion of this work was performed under the auspices of the US Department of Energy by Lawrence Livermore National Laboratory under contract DE-AC52-07NA27344. J.A.G. and G.J.A. were supported by LLNL Laboratory Directed Research and Development project 18-SI-002. The paper has LLNL tracking number LLNL-JRNL-811857. This document was prepared as an account of work sponsored by an agency of the United States government. Neither the United States government nor Lawrence Livermore National Security, LLC, nor any of their employees makes any warranty, expressed or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States government or Lawrence Livermore National Security, LLC. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States government or Lawrence Livermore National Security, LLC, and shall not be used for advertising or product endorsement purposes. Sandia National Laboratories is a multimission laboratory managed and operated by National Technology & Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International Inc., for the US Department of Energy's National Nuclear Security Administration under contract DE-NA0003525. This paper describes objective technical results and analysis. Any subjective views or opinions that might be expressed in the paper do not necessarily represent the views of the US Department of Energy or the United States Government.

## Author information

### Affiliations

1. Clarendon Laboratory, University of Oxford, Parks Road, Oxford, UK

Peter W. Hatfield & Steven Rose

2. Lawrence Livermore National Laboratory, Livermore, CA, USA

Jim A. Gaffney, Gemma J. Anderson, Suzanne Ali, Bogdan Kustowski, Michael J. MacDonald, Derek Mariscal, Madison E. Martin & J. Luc Peterson

3. York Plasma Institute, Department of Physics, University of York, York, UK

Luca Antonelli & Will Trickey

4. Nikhef, National Institute for Subatomic Physics, Amsterdam, The Netherlands

Suzan Başeğmez du Pree

5. DIFFER—Dutch Institute for Fundamental Energy Research, Eindhoven, The Netherlands

Jonathan Citrin

6. Instituto de Plasmas e Fusão Nuclear, Instituto Superior Técnico, Lisbon, Portugal

Marta Fajardo

7. Sandia National Laboratories, Albuquerque, NM, USA

Patrick Knapp & Taisuke Nagayama

8. Imperial College London, London, UK

Brendan Kettle, Steven Rose & Matt J. V. Streeter

9. School of Mathematics and Physics, Queen's University Belfast,  
Belfast, UK

Charlotte A. J. Palmer

10. Laboratory for Laser Energetics, University of Rochester, Rochester,  
NY, USA

J J Ruby

11. Dutch National Center for Mathematics and Computer Science (CWI),  
Amsterdam, The Netherlands

Carl Shneider

12. AWE Plc, Aldermaston, Reading, UK

Ben Williams

## Authors

1. Peter W. Hatfield

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

2. Jim A. Gaffney

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

3. Gemma J. Anderson

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

4. Suzanne Ali

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

5. Luca Antonelli

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

6. Suzan Başeğmez du Pree

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

7. Jonathan Citrin

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

8. Marta Fajardo

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

9. Patrick Knapp

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

10. Brendan Kettle

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

11. Bogdan Kustowski

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

12. Michael J. MacDonald

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

13. Derek Mariscal

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

14. Madison E. Martin

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

15. Taisuke Nagayama

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

16. Charlotte A. J. Palmer

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

17. J. Luc Peterson

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

18. Steven Rose

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

19. J J Ruby

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

20. Carl Shneider

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

21. Matt J. V. Streeter

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

22. Will Trickey

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

23. Ben Williams

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

## Contributions

P.W.H., J.A.G. and G.J.A. conceived the work and led the writing of the manuscript. All authors contributed to the manuscript and the ideas discussed at the Lorentz Center Meeting.

## Corresponding authors

Correspondence to [Peter W. Hatfield](#) or [Jim A. Gaffney](#) or [Gemma J. Anderson](#).

## Ethics declarations

## Competing interests

The authors declare no competing interests.

## Additional information

**Peer review information** *Nature* thanks Paul Bradley, Michael Bussmann and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

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

## Rights and permissions

### [Reprints and Permissions](#)

## About this article



Check for  
updates

## Cite this article

Hatfield, P.W., Gaffney, J.A., Anderson, G.J. *et al.* The data-driven future of high-energy-density physics. *Nature* **593**, 351–361 (2021).

<https://doi.org/10.1038/s41586-021-03382-w>

### [Download citation](#)

- Received: 24 June 2020
- Accepted: 22 February 2021
- Published: 19 May 2021
- Issue Date: 20 May 2021
- DOI: <https://doi.org/10.1038/s41586-021-03382-w>

## Comments

By submitting a comment you agree to abide by our [Terms](#) and [Community Guidelines](#). If you find something abusive or that does not comply with our terms or guidelines please flag it as inappropriate.

[Access through your institution](#)

[Change institution](#)

[Buy or subscribe](#)

Advertisement

---

This article was downloaded by **calibre** from <https://www.nature.com/articles/s41586-021-03382-w>

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

- Article
- [Published: 19 May 2021](#)

# Iron and nickel atoms in cometary atmospheres even far from the Sun

- [J. Manfroid<sup>1</sup>](#),
- [D. Hutsemékers<sup>1</sup>](#) &
- [E. Jehin<sup>1</sup>](#)

[Nature](#) volume 593, pages 372–374 (2021) [Cite this article](#)

- 752 Accesses
- 450 Altmetric
- [Metrics details](#)

## Subjects

- [Asteroids, comets and Kuiper belt](#)
- [Meteoritics](#)

## Abstract

In comets, iron and nickel are found in refractory dust particles or in metallic and sulfide grains<sup>1</sup>. So far, no iron- or nickel-bearing molecules have been observed in the gaseous coma of comets<sup>2</sup>. Iron and a few other heavy atoms, such as copper and cobalt, have been observed only in two exceptional objects: the Great Comet of 1882<sup>3</sup> and, almost a century later, C/1965 S1 (Ikeya–Seki)<sup>4,5,6,7,8,9</sup>. These sungrazing comets approached the Sun so closely that refractory materials sublimated, and their relative

abundance of nickel to iron was similar to that of the Sun and meteorites<sup>7</sup>. More recently, the presence of iron vapour was inferred from the properties of a faint tail in comet C/2006 P1 (McNaught) at perihelion<sup>10</sup>, but neither iron nor nickel was reported in the gaseous coma of comet 67P/Churyumov–Gerasimenko by the in situ Rosetta mission<sup>11</sup>. Here we report that neutral Fe i and Ni i emission lines are ubiquitous in cometary atmospheres, even far from the Sun, as revealed by high-resolution ultraviolet–optical spectra of a large sample of comets of various compositions and dynamical origins. The abundances of both species appear to be of the same order of magnitude, contrasting the typical Solar System abundance ratio.

## Access options

Subscribe to Journal

Get full journal access for 1 year

\$199.00

only \$3.90 per issue

[Subscribe](#)

All prices are NET prices.

VAT will be added later in the checkout.

Tax calculation will be finalised during checkout.

Rent or Buy article

Get time limited or full article access on ReadCube.

from \$8.99

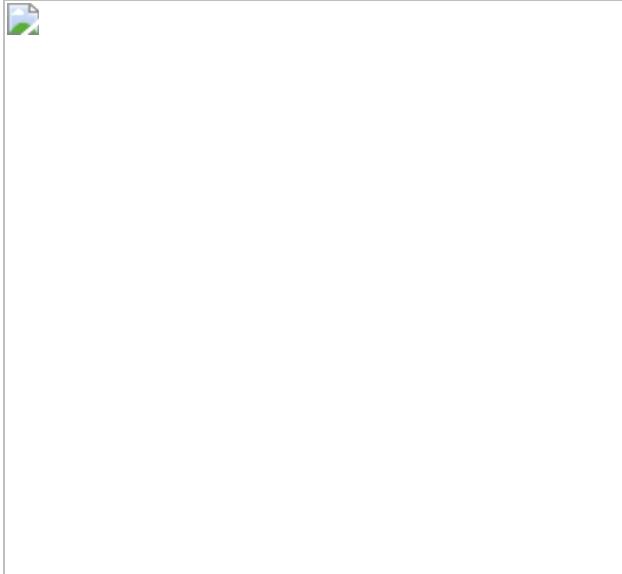
[Rent or Buy](#)

All prices are NET prices.

## **Additional access options:**

- [Log in](#)
- [Access through your institution](#)
- [Learn about institutional subscriptions](#)

**Fig. 1: Example of Fe i and Ni i lines in comet 103P/Hartley 2.**



**Fig. 2: Ni/Fe abundance ratios from the multilevel model versus heliocentric distance.**



## Data availability

The datasets analysed during the current study are available at the ESO Science Archive Facility at [http://archive.eso.org/eso/eso\\_archive\\_main.html](http://archive.eso.org/eso/eso_archive_main.html), under programme numbers 073.C-0525, 075.C- 0355(A), 080.C-0615, 086.C-0958, 087.C-0929, 270.C-5043, 274.C-5015, 2100.C-5035(A), 280.C-5053 and 2101.C-5051.

## References

1. 1.

Zolensky, M. E. et al. Mineralogy and petrology of comet 81P/Wild 2 nucleus samples. *Science* **314**, 1735–1739 (2006).

[ADS](#) [CAS](#) [Google Scholar](#)

2. 2.

Bockelée-Morvan, D. & Biver, N. The composition of cometary ices. *Philos. Trans. R. Soc. Lond. A* **375**, 20160252 (2017).

[ADS](#) [Google Scholar](#)

3. 3.

Copeland, R. & Lohse, J. G. in *Copernicus: an International Journal of Astronomy* Vol. 2, 225–244 (1882).

4. 4.

Arpigny, C. in *Liege International Astrophysical Colloquia* Vol. 22 (eds Boury, A. et al.) 189–197 (1979).

5. 5.

Combi, M. R., DiSanti, M. A. & Fink, U. The spatial distribution of gaseous atomic sodium in the comae of comets: evidence for direct nucleus and extended plasma sources. *Icarus* **130**, 336–354 (1997).

[ADS](#) [CAS](#) [Google Scholar](#)

6. 6.

Dufay, J., Swings, P. & Fehrenbach, Ch. Spectrographic observations of comet Ikeya-Seki (1965f). *Astrophys. J.* **142**, 1698 (1965).

[ADS](#) [CAS](#) [Google Scholar](#)

7. 7.

Preston, G. W. The spectrum of Ikeya-Seki (1965f). *Astrophys. J.* **147**, 718–742 (1967).

[ADS](#) [CAS](#) [Google Scholar](#)

8. 8.

Slaughter, C. D. The emission spectrum of comet Ikeya-Seki 1965-f at perihelion passage. *Astron. J.* **74**, 929 (1969).

[ADS](#) [Google Scholar](#)

9. 9.

Thackeray, A. D., Feast, M. W. & Warner, B. Daytime spectra of comet Ikeya-Seki near perihelion. *Astrophys. J.* **143**, 276 (1966).

[ADS](#) [CAS](#) [Google Scholar](#)

10. 10.

Fulle, M. et al. Discovery of the atomic iron tail of comet McNaught using the heliospheric imager on STEREO. *Astrophys. J. Lett.* **661**, L93–L96 (2007).

[ADS](#) [CAS](#) [Google Scholar](#)

11. 11.

Altwegg, K. & ROSINA Team. Chemical highlights from the Rosetta mission. In *Proceedings of the International Astronomical Union* Vol. 13, 153–162 (IAU, 2017).

12. 12.

Arpigny, C. et al. Anomalous nitrogen isotope ratio in comets. *Science* **301**, 1522–1524 (2003).

[ADS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

13. 13.

Guzik, P. & Drahus, M. Gaseous atomic nickel in the coma of interstellar comet 2I/Borisov. *Nature* <https://doi.org/10.1038/s41586-021-03485-4> (2021).

14. 14.

Arpigny, C. On the nature of comets. In *Proceedings of the Robert A. Welch Foundation Conferences on Chemical Research XXI, Cosmochemistry* (ed. Mulligan, W. O.) 9 (1978).

15. 15.

Cochran, A. L. & Schleicher, D. G. Observational constraints on the lifetime of cometary H<sub>2</sub>O. *Icarus* **105**, 235–253 (1993).

[ADS](#) [CAS](#) [Google Scholar](#)

16. 16.

Biver, N. et al. Long-term monitoring of the outgassing and composition of comet 67P/Churyumov–Gerasimenko with the Rosetta/MIRO instrument. *Astron. Astrophys.* **630**, A19 (2019).

[CAS](#) [Google Scholar](#)

17. 17.

Womack, M., Sarid, G. & Wierzchos, K. CO in distantly active comets. *Publ. Astron. Soc. Pacif.* **129**, 031001 (2017).

[ADS](#) [Google Scholar](#)

18. 18.

Lodders, K. Solar elemental abundances. In *Oxford Research Encyclopedia of Planetary Science* <https://doi.org/10.1093/acrefore/9780190647926.013.145> (2020).

19. 19.

Larimer, J. W. & Anders, E. Chemical fractionations in meteorites – III. Major element fractionations in chondrites. *Geochim. Cosmochim. Acta* **34**, 367–387 (1970).

[ADS](#) [CAS](#) [Google Scholar](#)

20. 20.

Grossman, L. & Larimer, J. W. Early chemical history of the solar system. *Rev. Geophys. Space Phys.* **12**, 71–101 (1974).

[ADS](#) [CAS](#) [Google Scholar](#)

21. 21.

Lewis, J. S. *Physics and Chemistry of the Solar System* 2nd edn (Academic Press, 2004).

22. 22.

Berger, E. L., Zega, T. J., Keller, L. P. & Lauretta, D. S. Evidence for aqueous activity on comet 81P/Wild 2 from sulfide mineral

assemblages in Stardust samples and CI chondrites. *Geochim. Cosmochim. Acta* **75**, 3501–3513 (2011).

[ADS](#) [CAS](#) [Google Scholar](#)

23. 23.

Bradley, J. P. Chemically anomalous, preaccretionaly irradiated grains in interplanetary dust from comets. *Science* **265**, 925–929 (1994).

[ADS](#) [CAS](#) [Google Scholar](#)

24. 24.

Ishii, H. A. et al. Comparison of comet 81P/Wild 2 dust with interplanetary dust from comets. *Science* **319**, 447 (2008).

[ADS](#) [CAS](#) [Google Scholar](#)

25. 25.

Bockelée-Morvan, D. et al. Comet 67P outbursts and quiescent coma at 1.3 au from the Sun: dust properties from Rosetta/VIRTIS-H observations. *Mon. Not. R. Astron. Soc.* **469**, S443–S458 (2017).

[Google Scholar](#)

26. 26.

Hensley, B. S. & Draine, B. T. Thermodynamics and charging of interstellar iron nanoparticles. *Astrophys. J.* **834**, 134 (2017).

[ADS](#) [Google Scholar](#)

27. 27.

Ip, W.-H. & Jorda, L. Can the sodium tail of Comet Hale–Bopp have a dust-impact origin? *Astrophys. J.* **496**, L47–L49 (1998).

[ADS](#) [CAS](#) [Google Scholar](#)

28. 28.

Wurz, P. et al. Solar wind sputtering of dust on the surface of 67P/Churyumov–Gerasimenko. *Astron. Astrophys.* **583**, A22 (2015).

[Google Scholar](#)

29. 29.

Bloch, M. R. & Wirth, H. L. Abiotic organic synthesis in space. *Naturwissenschaften* **67**, 562–564 (1980).

[ADS](#) [CAS](#) [Google Scholar](#)

30. 30.

Klotz, A. et al. Possible contribution of organometallic species in the solar system ices. Reactivity and spectroscopy. *Planet. Space Sci.* **44**, 957–965 (1996).

[ADS](#) [CAS](#) [Google Scholar](#)

31. 31.

Huebner, W. F. Dust from cometary nuclei. *Astron. Astrophys.* **5**, 286–297 (1970).

[ADS](#) [CAS](#) [Google Scholar](#)

32. 32.

Tarakeshwar, P., Buseck, P. R. & Timmes, F. X. On the structure, magnetic properties, and infrared spectra of iron pseudocarbynes in the interstellar medium. *Astrophys. J.* **879**, 2 (2019).

[ADS](#) [CAS](#) [Google Scholar](#)

33. 33.
- Smith, K. E., House, C. H., Arevalo, R. D., Dworkin, J. P. & Callahan, M. P. Organometallic compounds as carriers of extraterrestrial cyanide in primitive meteorites. *Nat. Commun.* **10**, 2777 (2019).

[ADS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

34. 34.
- Prialnik, D., Benkhoff, J. & Podolak, M. in *Comets II* (eds Festou et al.) 359–387 (Univ. of Arizona Press, 2004).

35. 35.
- Levison, H. F. Comet taxonomy. In *Completing the Inventory of the Solar System* Vol. 107 (eds. Rettig, T. W. & Hahn, J. M.) 173–191 (ASP, 1996).

36. 36.
- A'Hearn, M. F., Schleicher, D. G., Millis, R. L., Feldman, P. D. & Thompson, D. T. Comet Bowell 1980b. *Astron. J.* **89**, 579–591 (1984).

[ADS](#) [Google Scholar](#)

37. 37.
- Schleicher, D. G. The fluorescence efficiencies of the CN violet bands in comets. *Astron. J.* **140**, 973–984 (2010).

[ADS](#) [CAS](#) [Google Scholar](#)

38. 38.
- Schleicher, D. G. & A'Hearn, M. F. The fluorescence of cometary OH. *Astrophys. J.* **331**, 1058 (1988).

[ADS](#) [CAS](#) [Google Scholar](#)

39. 39.

Bhardwaj, A. & Raghuram, S. A coupled chemistry-emission model for atomic oxygen green and red-doublet emissions in the comet C/1996 B2 Hyakutake. *Astrophys. J.* **748**, 13 (2012).

[ADS](#) [Google Scholar](#)

40. 40.

Raghuram, S. et al. A physico-chemical model to study the ion density distribution in the inner coma of comet C/2016 R2(Pan-STARRS). *Mon. Not. R. Astron. Soc.* **501**, 4035–4052 (2021).

41. 41.

Weaver, H., Feldman, P., A'Hearn, M., Dello Russo, N. & Stern, A. Comet 103P/Hartley. *IAU Circ.* **9183**, 1 (2010).

[ADS](#) [Google Scholar](#)

42. 42.

Roth, N. X. et al. Probing the evolutionary history of comets: an investigation of the hypervolatiles CO, CH<sub>4</sub>, and C<sub>2</sub>H<sub>6</sub> in the Jupiter-family Comet 21P/Giacobini–Zinner. *Astron. J.* **159**, 42 (2020).

[ADS](#) [CAS](#) [Google Scholar](#)

43. 43.

DiSanti, M. A. et al. Depleted carbon monoxide in fragment C of the Jupiter-family comet 73P/Schwassmann–Wachmann 3. *Astrophys. J. Lett.* **661**, L101–L104 (2007).

[ADS](#) [CAS](#) [Google Scholar](#)

44. 44.

Böhnhardt, H. et al. The unusual volatile composition of the Halley-type comet 8P/Tuttle: addressing the existence of an inner Oort cloud. *Astrophys. J. Lett.* **683**, L71 (2008).

[ADS](#) [Google Scholar](#)

45. 45.

Lupu, R. E., Feldman, P. D., Weaver, H. A. & Tozzi, G.-P. The fourth positive system of carbon monoxide in the Hubble Space Telescope spectra of comets. *Astrophys. J.* **670**, 1473–1484 (2007).

[ADS](#) [CAS](#) [Google Scholar](#)

46. 46.

DiSanti, M. A. et al. Detection of formaldehyde emission in comet C/2002 T7 (LINEAR) at infrared wavelengths: line-by-line validation of modeled fluorescent intensities. *Astrophys. J.* **650**, 470–483 (2006).

[ADS](#) [CAS](#) [Google Scholar](#)

47. 47.

Paganini, L. et al. Observations of comet C/2009 P1 (Garradd) at 2.4 and 2.0 AU before perihelion. In *Asteroids, Comets, Meteors 2012* Vol. 1667, 6331 (Lunar Planetary Institute, 2012).

48. 48.

Paganini, L. et al. The unexpectedly bright comet C/2012 F6 (Lemmon) unveiled at near-infrared wavelengths. *Astron. J.* **147**, 15 (2013).

[ADS](#) [Google Scholar](#)

49. 49.

Biver, N. et al. The extraordinary composition of the blue comet C/2016 R2 (PanSTARRS). *Astron. Astrophys.* **619**, A127 (2018).

[ADS](#) [CAS](#) [Google Scholar](#)

50. 50.

Wierzchos, K. & Womack, M. C/2016 R2 (PANSTARRS): a comet rich in CO and depleted in HCN. *Astron. J.* **156**, 34 (2018).

[ADS](#) [CAS](#) [Google Scholar](#)

51. 51.

Faggi, S., Mumma, M. J., Villanueva, G. L., Paganini, L. & Lippi, M. Quantifying the evolution of molecular production rates of comet 21P/Giacobini–Zinner with iSHELL/NASA-IRTF. *Astron. J.* **158**, 254 (2019).

[ADS](#) [CAS](#) [Google Scholar](#)

52. 52.

de Val-Borro, M. et al. A survey of volatile species in Oort cloud comets C/2001 Q4 (NEAT) and C/2002 T7 (LINEAR) at millimeter wavelengths. *Astron. Astrophys.* **559**, A48 (2013).

[Google Scholar](#)

53. 53.

Yamamoto, T., Nakagawa, N. & Fukui, Y. The chemical composition and thermal history of the ice of a cometary nucleus. *Astron. Astrophys.* **122**, 171–176 (1983).

[ADS](#) [CAS](#) [Google Scholar](#)

54. 54.

Yamamoto, T. Formation environment of cometary nuclei in the primordial solar nebula. *Astron. Astrophys.* **142**, 31–36 (1985).

[ADS](#) [CAS](#) [Google Scholar](#)

55. 55.

Gilbert, A. G. & Sulzmann, K. G. P. The vapor pressure of iron pentacarbonyl. *J. Electrochem. Soc.* **121**, 832 (1974).

[ADS](#) [CAS](#) [Google Scholar](#)

56. 56.

Stull, D. R. Vapor pressure of pure substances. Organic and inorganic compounds. *Ind. Eng. Chem.* **39**, 517–540 (1947).

[CAS](#) [Google Scholar](#)

57. 57.

Delsemme, A. H. Chemical composition of cometary nuclei. In *IAU Colloq. 61: Comet Discoveries, Statistics, and Observational Selection* (ed. Wilkening, L. L.) 85–130 (IAU, 1982).

[Download references](#)

## Acknowledgements

We thank P. van Hoof for discussions on the iron atomic data and their uncertainties. We thank R. Hewins and R. Warin for discussions about various Fe- and Ni-rich compounds in meteorites. We thank C. Arpigny, D. Bockelée-Morvan, A. Decock, C. Opitom, H. Rauer, P. Rousselot and B. Yang for leading some UVES proposals, and the ESO staff for service mode observations. J.M., D.H. and E.J. are Honorary Research Director, Research Director and Senior Research Associate at the Fonds de la Recherche Scientifique (F.R.S-FNRS), respectively.

# Author information

## Affiliations

1. STAR Institute, University of Liège, Liège, Belgium

J. Manfroid, D. Hutsemékers & E. Jehin

## Authors

1. J. Manfroid

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

2. D. Hutsemékers

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

3. E. Jehin

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

## Contributions

J.M. analysed the spectra and the coma profiles and wrote the main text. D.H. contributed to the proposals and observations, reduced and calibrated the spectra, built the fluorescence model, computed the carbonyl sublimation properties and wrote the Supplementary Information. E.J. led the UVES proposals and made most of the observations. All authors contributed to the discussion and the final text.

## Corresponding author

Correspondence to [J. Manfroid](#).

# Ethics declarations

## Competing interests

The authors declare no competing interests.

## Additional information

**Peer review information** *Nature* thanks Dennis Bodewits and Ryan Fortenberry for their contribution to the peer review of this work.

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

## Extended data figures and tables

### [Extended Data Fig. 1 Examples of UVES comet spectra.](#)

Comet spectra obtained with the UVES spectrograph at ESO VLT, showing many Fe i and Ni i lines in the selected wavelength region (3,425–3,530 Å). **a**, Spectrum of the water-poor and CO-rich long-period comet C/2016 R2 (PanSTARRS) at 3 au. **b**, Spectrum of the Jupiter-family comet 88P/Howell at 1.4 au. **c**, Spectrum of the new comet C/2002 T7 (LINEAR) at 0.68 au, with lines from the OH(1-2) band. **d**, Spectrum of the long-period comet C/2020 X5 (Kudo–Fujikawa). Fe i and Ni i lines are indicated by red and blue marks, respectively.

### [Extended Data Fig. 2 Comparisons of Fe i, Ni i and dust production rates.](#)

**a–d**, The production rates of Fe i and Ni i are compared to  $Afp$  (which is the product of the reflectivity of the grains, their filling factor and the radius of the coma; used as a proxy for the dust production rate) and to the production rates of OH, CN and  $\text{CO}_2^+$ , as determined from our spectra. **e–f**, The production rates of Fe i and Ni i are compared to those of  $\text{H}_2\text{O}$  and CO

measured in various previous studies for comets 8P, 9P, 21P, 73P, 103P, C/2000 WM1, C/2001 Q4, C/2002 T7, C/2009 P1, C/2012 F6 and C/2016 R2 at about the same epochs as our observations<sup>41,42,43,44,45,46,47,48,49,50,51,52</sup>. The various cometary types are colour-coded according to their dynamical classification (see Extended Data Table 1). The OH and H<sub>2</sub>O values relative to comet C/2016 R2 are upper limits. The Pearson correlation coefficients calculated without (with) the C/2016 R2 data are  $\rho_{\text{OH}} = 0.844$  (0.531),  $\rho_{\text{Afp}} = 0.644$  (0.616),  $\rho_{\text{CN}} = 0.892$  (0.518),  $\langle \{\rho\}_{\{\{\{\rm CO\}\}\{2\}\}^{\{+\}}}\rangle = 0.755$  (0.804),  $\langle \{\rho\}_{\{\{\{\rm H\}\}\}\{2\}\{\rm O\}}\rangle = 0.849$  (0.627) and  $\rho_{\text{CO}} = 0.752$  (0.770).

### Extended Data Fig. 3 Iron and nickel carbonyl sublimation properties.

**a**, Sublimation rates ( $Z$ ; in molecules cm<sup>-2</sup> s<sup>-1</sup>) of Fe and Ni carbonyls as a function of temperature, compared to those of the main ices in comets. The carbonyl rates are intermediate between those of H<sub>2</sub>O and CO<sub>2</sub>. **b**, The ratio of the sublimation rate of Ni(CO)<sub>4</sub> over that of Fe(CO)<sub>5</sub> shows that the former is considerably higher than the latter. These quantities were computed as follows. As in refs. <sup>53,54</sup>, we estimate the condensation or sublimation temperature  $T_s$  of these compounds by solving the equation  $f_x n k T_s = P_{v,x}(T_s)$  where  $f_x$  is the relative abundance of species  $x$ ,  $n$  is the number density of the gas,  $k$  is the Boltzmann constant, and  $P_{v,x}$  is the vapour pressure, given by the relation  $\log[P_{v,x}(T)] = -(A/T) + B$ . The constants  $A$  and  $B$  for Fe(CO)<sub>5</sub> and Ni(CO)<sub>4</sub> are obtained from refs. <sup>55,56</sup>:  $A = 2,097$  K and  $B = 11.62$  for Fe(CO)<sub>5</sub>,  $A = 1,534$  K and  $B = 10.87$  for Ni(CO)<sub>4</sub>, with  $P_{v,x}$  in dyn cm<sup>-2</sup>. We consider relative abundances  $f_x$  of  $10^{-3}$ – $10^{-5} \times f_x(\text{H}_2\text{O})$  for both Fe(CO)<sub>5</sub> and Ni(CO)<sub>4</sub>, and we adopt  $n = 10^{13}$  cm<sup>-3</sup> as in ref. <sup>54</sup>. The resulting sublimation temperatures of the iron and nickel carbonyls (97–108 K and 74–82 K, respectively, depending on  $f_x$ ) are between the sublimation temperatures of H<sub>2</sub>O and CO<sub>2</sub> (152 K and 72 K), whereas CO sublimates at 25 K (ref. <sup>54</sup>). The sublimation rate (in molecules cm<sup>-2</sup> s<sup>-1</sup>) from the surface of a pure ice into vacuum can be expressed as<sup>57</sup>:

$Z_x(T) = P_{v,x}(T)(2\pi m_x kT)^{-1/2}$ , where  $T$  is the ice temperature and  $m_x$  the mass of species  $x$ .

**Extended Data Table 1 Conditions of comet observations with UVES at ESO VLT**

[Full size table](#)

**Extended Data Table 2 Ni/Fe abundance ratios from the three-level and multilevel models**

[Full size table](#)

**Extended Data Table 3 Production rates of molecules and dust**

[Full size table](#)

## Supplementary information

### [Supplementary Information](#)

This file contains details regarding the FeI and NiI fluorescence models and abundance measurements, Supplementary Figures 1–6 and Supplementary References.

## Rights and permissions

[Reprints and Permissions](#)

## About this article



Check for  
updates

## Cite this article

Manfroid, J., Hutsemékers, D. & Jehin, E. Iron and nickel atoms in cometary atmospheres even far from the Sun. *Nature* **593**, 372–374 (2021). <https://doi.org/10.1038/s41586-021-03435-0>

## Download citation

- Received: 01 November 2020
- Accepted: 09 March 2021
- Published: 19 May 2021
- Issue Date: 20 May 2021
- DOI: <https://doi.org/10.1038/s41586-021-03435-0>

## Comments

By submitting a comment you agree to abide by our [Terms](#) and [Community Guidelines](#). If you find something abusive or that does not comply with our terms or guidelines please flag it as inappropriate.

[Access through your institution](#)

[Change institution](#)

[Buy or subscribe](#)

## Associated Content

### [\*\*Iron and nickel vapours are present in most comets\*\*](#)

- Dennis Bodewits
- Steven J. Bromley

Nature News & Views 19 May 2021

Advertisement

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

- Article
- [Published: 19 May 2021](#)

# Gaseous atomic nickel in the coma of interstellar comet 2I/Borisov

- [Piotr Guzik](#) [ORCID: orcid.org/0000-0002-5019-4056<sup>1</sup>](#) &
- [Michał Drahus](#) [ORCID: orcid.org/0000-0001-5159-2104<sup>1</sup>](#)

[Nature](#) volume 593, pages 375–378 (2021) [Cite this article](#)

- 444 Accesses
- 1 Citations
- 515 Altmetric
- [Metrics details](#)

## Subjects

- [Asteroids, comets and Kuiper belt](#)
- [Exoplanets](#)

## Abstract

On 31 August 2019, an interstellar comet was discovered as it passed through the Solar System (2I/Borisov). On the basis of initial imaging observations, 2I/Borisov seemed to be similar to ordinary Solar System comets<sup>1,2</sup>—an unexpected characteristic given the multiple peculiarities of the only known previous interstellar visitor, 1I/‘Oumuamua<sup>3,4,5,6</sup>. Spectroscopic investigations of 2I/Borisov identified the familiar cometary

emissions from CN (refs. [7,8,9](#)), C<sub>2</sub> (ref. [10](#)), O i (ref. [11](#)), NH<sub>2</sub> (ref. [12](#)), OH (ref. [13](#)), HCN (ref. [14](#)) and CO (refs. [14,15](#)), revealing a composition similar to that of carbon monoxide-rich Solar System comets. At temperatures greater than 700 kelvin, comets also show metallic vapours that are produced by the sublimation of metal-rich dust grains<sup>[16](#)</sup>. Observation of gaseous metals had until very recently<sup>[17](#)</sup> been limited to bright sunskirting and sungrazing comets<sup>[18,19,20](#)</sup> and giant star-plunging exocomets<sup>[21](#)</sup>. Here we report spectroscopic observations of atomic nickel vapour in the cold coma of 2I/Borisov at a heliocentric distance of 2.322 astronomical units—equivalent to an equilibrium temperature of 180 kelvin. Nickel in 2I/Borisov seems to originate from a short-lived nickel-containing molecule with a lifetime of  $\sqrt{340 - 200 + 260}$  seconds at 1 astronomical unit and is produced at a rate of  $0.9 \pm 0.3 \times 10^{22}$  atoms per second, or 0.002 per cent relative to OH and 0.3 per cent relative to CN. The detection of gas-phase nickel in the coma of 2I/Borisov is in line with the recent identification of this atom—as well as iron—in the cold comae of Solar System comets<sup>[17](#)</sup>.

## Access options

Subscribe to Journal

Get full journal access for 1 year

\$199.00

only \$3.90 per issue

[Subscribe](#)

All prices are NET prices.

VAT will be added later in the checkout.

Tax calculation will be finalised during checkout.

Rent or Buy article

Get time limited or full article access on ReadCube.

from \$8.99

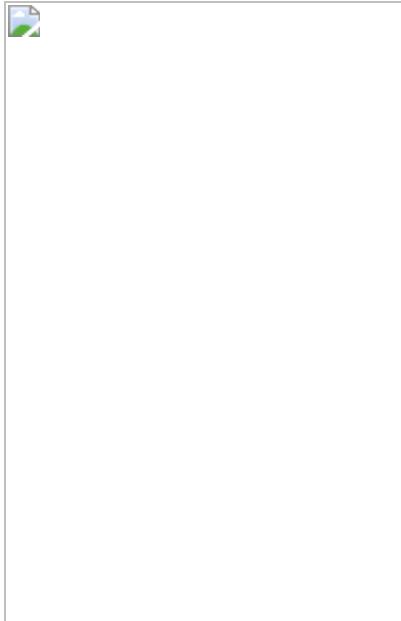
[Rent or Buy](#)

All prices are NET prices.

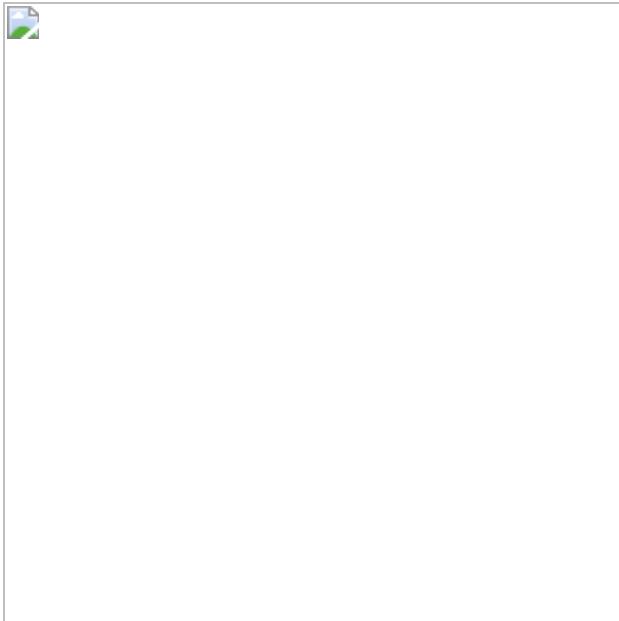
### **Additional access options:**

- [Log in](#)
- [Access through your institution](#)
- [Learn about institutional subscriptions](#)

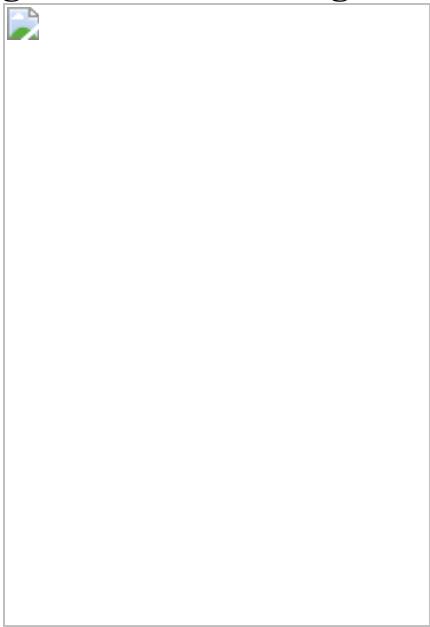
**Fig. 1: Emission lines from gaseous atomic nickel in the near-UV spectrum of 2I/Borisov.**



**Fig. 2: Observed and modelled spatial profiles of nickel emission.**



**Fig. 3: Haser scalelengths of the observed nickel emission.**



## Data availability

The X-shooter raw data are available in the ESO archive at  
<https://archive.eso.org>. [Source data](#) are provided with this paper.

## Code availability

The EsoReflex pipeline is available from the ESO website at <https://www.eso.org/sci/software/esoreflex/>. All custom codes are direct implementations of standard methods and techniques, described in detail in Methods.

## References

1. 1.

Guzik, P. et al. Initial characterization of interstellar comet 2I/Borisov. *Nat. Astron.* **4**, 53–57 (2020).

[ADS](#) [Google Scholar](#)

2. 2.

Jewitt, D. & Luu, J. Initial characterization of interstellar comet 2I/2019 Q4 (Borisov). *Astrophys. J. Lett.* **886**, 29 (2019).

3. 3.

Fraser, W. C. et al. The tumbling rotational state of 1I/‘Oumuamua. *Nat. Astron.* **2**, 383–386 (2018).

[ADS](#) [Google Scholar](#)

4. 4.

Meech, K. J. et al. A brief visit from a red and extremely elongated interstellar asteroid. *Nature* **552**, 378–381 (2017).

[ADS](#) [CAS](#) [Google Scholar](#)

5. 5.

Drahus, M. et al. Tumbling motion of 1I/‘Oumuamua and its implications for the body’s distant past. *Nat. Astron.* **2**, 407–412 (2018).

[ADS](#) [Google Scholar](#)

6. 6.

Micheli, M. et al. Non-gravitational acceleration in the trajectory of 1I/2017 U1 ('Oumuamua). *Nature* **559**, 223–226 (2018).

[ADS](#) [CAS](#) [Google Scholar](#)

7. 7.

Fitzsimmons, A. et al. Detection of CN gas in interstellar object 2I/Borisov. *Astrophys. J.* **885**, L9 (2019).

[ADS](#) [CAS](#) [Google Scholar](#)

8. 8.

De León, J. et al. Visible and near-infrared observations of interstellar comet 2I/Borisov with the 10.4-m GTC and the 3.6-m TNG telescopes. *Mon. Not. R. Astron. Soc.* **495**, 2053–2062 (2020).

[ADS](#) [Google Scholar](#)

9. 9.

Kareta, T. et al. Carbon chain depletion of 2I/Borisov. *Astrophys. J.* **889**, L38 (2020).

[ADS](#) [CAS](#) [Google Scholar](#)

10. 10.

Lin, H. W. et al. Detection of diatomic carbon in 2I/Borisov. *Astrophys. J.* **889**, L30 (2020).

[ADS](#) [CAS](#) [Google Scholar](#)

11. 11.

McKay, A. J., Cochran, A. L., Dello Russo, N. & DiSanti, M. A. Detection of a water tracer in interstellar comet 2I/Borisov. *Astrophys. J.* **889**, L10 (2020).

[ADS](#) [CAS](#) [Google Scholar](#)

12. 12.

Bannister, M. T. et al. Interstellar comet 2I/Borisov as seen by MUSE: C2, NH<sub>2</sub> and red CN detections. Preprint at <https://arxiv.org/abs/2001.11605> (2020).

13. 13.

Jehin, E. et al. CBET 4719: COMET 2I/2019 Q4. <http://www.cbat.eps.harvard.edu/cbet/004700/CBET004719.txt> (2020).

14. 14.

Cordiner, M. A. et al. Unusually high CO abundance of the first active interstellar comet. *Nat. Astron.* **4**, 861–866 (2020).

[ADS](#) [Google Scholar](#)

15. 15.

Bodewits, D. et al. The carbon monoxide-rich interstellar comet 2I/Borisov. *Nat. Astron.* **4**, 867–871 (2020).

[ADS](#) [Google Scholar](#)

16. 16.

Feldman, P., Cochran, A. & Combi, M. in *Comets II* (eds Festou, M., Keller, H. U. & Weaver, H. A.) 425–447 (Univ. Arizona Press, 2004).

17. 17.

Manfroid, J., Hutsemékers, D. & Jehin, E. Iron and nickel atoms in cometary atmospheres even far from the Sun. *Nature* <https://doi.org/10.1038/s41586-021-03435-0> (2021).

18. 18.

Preston, G. The spectrum of comet Ikeya-Seki (1965f). *Astrophys. J.* **147**, 718–742 (1967).

[ADS](#) [CAS](#) [Google Scholar](#)

19. 19.

Slaughter, C. The emission spectrum of comet Ikeya-Seki 1965-f at perihelion passage. *Astron. J.* **74**, 929–943 (1969).

[ADS](#) [Google Scholar](#)

20. 20.

Fulle, M. et al. Discovery of the atomic iron tail of comet McNaught using the heliospheric imager on STEREO. *Astrophys. J.* **661**, L93 (2007).

[ADS](#) [CAS](#) [Google Scholar](#)

21. 21.

Kiefer, F. et al. Fe I in the  $\beta$  Pictoris circumstellar gas disk. II. Time variations in the iron circumstellar gas. *Astron. Astrophys.* **621**, A58 (2019).

[CAS](#) [Google Scholar](#)

22. 22.

Kim, S. J., A'Hearn, M. F., Wellnitz, D. D., Meier, R. & Lee, Y. S. The rotational structure of the B–X system of sulfur dimers in the spectra of Comet Hyakutake (C/1996 B2). *Icarus* **166**, 157–166 (2003).

23. 23.

Valk, J. H. & O'Dell, C. R. Near-ultraviolet spectroscopy of comet Austin (1989c1). *Astrophys. J.* **388**, 621–632 (1992).

[ADS](#) [CAS](#) [Google Scholar](#)

24. 24.

Cochran, A. L. A search for N<sub>2</sub><sup>+</sup> in spectra of comet C/2002 C1 (Ikeya-Zhang). *Astrophys. J.* **576**, L165–L168 (2002).

[ADS](#) [CAS](#) [Google Scholar](#)

25. 25.

Cochran, A. L. & McKay, A. J. Strong CO<sup>+</sup> and N<sub>2</sub><sup>+</sup> emission in comet C/2016 R2 (Pan-STARRS). *Astrophys. J. Lett.* **854**, L10 (2018).

[ADS](#) [Google Scholar](#)

26. 26.

Opitom, C. et al. High resolution optical spectroscopy of the N<sub>2</sub>-rich comet C/2016 R2 (PanSTARRS). *Astron. Astrophys.* **624**, A64 (2019).

[CAS](#) [Google Scholar](#)

27. 27.

Swings, P. Complex structure of cometary bands tentatively ascribed to the contour of the solar spectrum. *Lick Obs. Bull.* **19**, 131 (1941).

[ADS](#) [CAS](#) [Google Scholar](#)

28. 28.

Peck, E. R. & Reeder, K. Dispersion of air. *J. Opt. Soc. Am.* **62**, 958–962 (1972).

29. 29.

Haser, L. Distribution d'intensité dans la tête d'une comète. *Bull. Acad. R. Sci. Liege* **43**, 740–750 (1957).

[ADS](#) [MathSciNet](#) [MATH](#) [Google Scholar](#)

30. 30.

Combi, M. R., Harris, W. M. & Skyth, W. M. in *Comets II* (eds Festou, M., Keller, H. U. & Weaver, H. A.) 523–554 (Univ. Arizona Press, 2004).

31. 31.

Huebner, W. F. & Mukherjee, J. Photoionization and photodissociation rates in solar and blackbody radiation fields. *Planet. Space Sci.* **106**, 11–45 (2015).

[ADS](#) [CAS](#) [Google Scholar](#)

32. 32.

A'Hearn, M. F., Millis, R. C., Schleicher, D. G., Osip, D. J. & Birch, P. V. The ensemble properties of comets: results from narrowband photometry of 85 comets, 1976–1992. *Icarus* **118**, 223–270 (1995).

[ADS](#) [Google Scholar](#)

33. 33.

Johnson, J. A. Populating the periodic table: nucleosynthesis of the elements. *Science* **363**, 474–478 (2019).

[ADS](#) [CAS](#) [Google Scholar](#)

34. 34.

Wasson, J. T. *Meteorites* 11–38 (Springer, 1974).

35. 35.

Goldberg, R. A. & Aikin, A. C. Comet Encke: meteor metallic ion identification by mass spectrometer. *Science* **180**, 294–296 (1973).

[ADS](#) [CAS](#) [Google Scholar](#)

36. 36.

Jessberger, E. K., Christoforidis, A. & Kissel, J. Aspects of the major element composition of Halley's dust. *Nature* **332**, 691–695 (1988).

[ADS](#) [CAS](#) [Google Scholar](#)

37. 37.

Flynn, G. J. et al. Elemental compositions of comet 81P/Wild 2 samples collected by Stardust. *Science* **314**, 1731–1735 (2006).

[ADS](#) [CAS](#) [Google Scholar](#)

38. 38.

Zolensky, M. E. et al. Mineralogy and petrology of comet 81P/Wild 2 nucleus samples. *Science* **314**, 1735–1739 (2006).

[ADS](#) [CAS](#) [Google Scholar](#)

39. 39.

Chochol, D., Rušín, V., Kulčár, L. & Vanýsek, V. Emission features in the solar corona after the perihelion passage of Comet 1979 XI. *Astrophys. Space Sci.* **91**, 71–77 (1983).

[ADS](#) [CAS](#) [Google Scholar](#)

40. 40.

Hoeijmakers, H. J. et al. Atomic iron and titanium in the atmosphere of the exoplanet KELT-9b. *Nature* **560**, 453–455 (2018).

[ADS](#) [CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

41. 41.

Hoeijmakers, H. J. et al. Hot exoplanet atmospheres resolved with transit spectroscopy (HEARTS). *Astron. Astrophys.* **641**, A123 (2020).

[CAS](#) [Google Scholar](#)

42. 42.

Vanderburg, A. et al. A disintegrating minor planet transiting a white dwarf. *Nature* **526**, 546–549 (2015).

[ADS](#) [CAS](#) [Google Scholar](#)

43. 43.

Xu, S. et al. The chemical composition of an extrasolar Kuiper-belt-object. *Astrophys. J. Lett.* **836**, L7 (2017).

[ADS](#) [Google Scholar](#)

44. 44.

Prialnik, D., Benkhoff, J. & Podolak, M. in *Comets II* (eds Festou, M., Keller, H. U. & Weaver, H. A.) 359–387 (Univ. Arizona Press, 2004).

45. 45.

Rubin, M. et al. Elemental and molecular abundances in comet 67P/Churyumov–Gerasimenko. *Mon. Not. R. Astron. Soc.* **489**, 594–607 (2019).

[ADS](#) [CAS](#) [Google Scholar](#)

46. 46.

Ivezić, Ž., et al. LSST: from science drivers to reference design and anticipated data products. *Astrophys. J.* **873**, 111 (2019).

[ADS](#) [Google Scholar](#)

47. 47.

Freudling, W. et al. Automated data reduction workflows for astronomy. The ESO Reflex environment. *Astron. Astrophys.* **559**, A96 (2013).

[Google Scholar](#)

48. 48.

Kurucz, R. L. New atlases for solar flux, irradiance, central intensity, and limb intensity. *Mem. Soc. Astron. Ital.* **8**, 189 (2005).

[ADS](#) [Google Scholar](#)

49. 49.

Kramida, A., Ralchenko, Yu., Reader, J. & NIST ASD Team. *NIST Atomic Spectra Database (v.5.7.1)* (National Institute of Standards and Technology, accessed 22 October 2020);  
<https://doi.org/10.18434/T4W30F>.

50. 50.

Haser, L., Oset, S. & Bodewits, D. Intensity distribution in the heads of comets. *Planet. Sci. J.* **1**, 83 (2020).

[Google Scholar](#)

51. 51.

Combi, M. R. & Delsemme, A. H. Neutral cometary atmospheres. I—An average random walk model for photodissociation in comets. *Astrophys. J.* **237**, 633–640 (1980).

[ADS](#) [CAS](#) [Google Scholar](#)

52. 52.

Schleicher, D. G. & A'Hearn, M. F. The fluorescence of cometary OH. *Astrophys. J.* **331**, 1058–1077 (1988).

[ADS](#) [CAS](#) [Google Scholar](#)

53. 53.

Schleicher, D. G. The fluorescence efficiencies of the CN violet bands in comets. *Astron. J.* **140**, 973–984 (2010).

[ADS](#) [CAS](#) [Google Scholar](#)

54. 54.

Bohlin, R. C., Gordon, K. D. & Tremblay P.-E. Techniques and review of absolute flux calibration from the ultraviolet to the mid-infrared. *Publ. Astron. Soc. Pac.* **126**, 711 (2014).

55. 55.

Hall, L. A. & Anderson, G. P. High-resolution solar spectrum between 2000 and 3100 Å. *J. Geophys. Res.* **96**, 12927 (1991).

[Download references](#)

## Acknowledgements

We thank K. Rusek for help with proposal writing, M. Ratajczak and M. Gromadzki for introducing us to X-shooter data reduction, and P. Kozyra for discussion on nickel-containing molecules. This work is based on

observations collected at the ESO under ESO programme 0104.C-0933(B). We thank the ESO staff for support. We are also grateful for support from the National Science Centre of Poland through ETIUDA scholarship no. 2020/36/T/ST9/00596 to P.G. and SONATA BIS grant no. 2016/22/E/ST9/00109 to M.D., and we acknowledge support from the Polish Ministry of Science and Higher Education through grant no. DIR/WK/2018/12.

## Author information

### Affiliations

1. Astronomical Observatory, Jagiellonian University, Kraków, Poland

Piotr Guzik & Michał Drahus

### Authors

1. Piotr Guzik

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

2. Michał Drahus

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

### Contributions

P.G. and M.D. wrote the telescope time proposal, searched for the origin of the detected spectral lines and wrote the paper. P.G. prepared the observations, reduced and calibrated the data, identified the emitting species and measured the spectral lines. M.D. created the fluorescence model, retrieved the scalelengths and calculated the production rate.

### Corresponding authors

Correspondence to [Piotr Guzik](#) or [Michał Drahus](#).

## Ethics declarations

### Competing interests

The authors declare no competing interests.

## Additional information

**Peer review information** *Nature* thanks Ryan Fortenberry and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

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

## Extended data figures and tables

### [Extended Data Fig. 1 Complete spectrum of comet 2I/Borisov from X-shooter UVB arm.](#)

**a**, Flux-calibrated spectrum with fitted dust continuum (see [Methods](#)). **b**, Same as **a** but with the dust-continuum component removed. Major emission features are labelled. **c**, Modelled spectrum of nickel fluorescence emission (see [Methods](#)) scaled to best match the two brightest lines.

### [Extended Data Fig. 2 Distribution of Monte Carlo-simulated production rates.](#)

The distribution was constructed from the production rates corresponding to the results of the Monte Carlo simulation in Fig. [3b](#) (see [Methods](#)). Results are presented in three groups according to the assumed PSF equal to 0.65 (blue), 1.0 (red) and 1.5 (green) arcsec.

# Source data

[Source Data Fig. 1](#)

[Source Data Fig. 2](#)

[Source Data Fig. 3](#)

# Rights and permissions

[Reprints and Permissions](#)

# About this article



Check for  
updates

# Cite this article

Guzik, P., Drahus, M. Gaseous atomic nickel in the coma of interstellar comet 2I/Borisov. *Nature* **593**, 375–378 (2021).  
<https://doi.org/10.1038/s41586-021-03485-4>

[Download citation](#)

- Received: 24 October 2020
- Accepted: 23 March 2021
- Published: 19 May 2021
- Issue Date: 20 May 2021
- DOI: <https://doi.org/10.1038/s41586-021-03485-4>

## Further reading

- [Iron and nickel atoms in cometary atmospheres even far from the Sun](#)
  - J. Manfroid
  - , D. HutsemÃ©kers
  - & E. Jehin

*Nature* (2021)

## Comments

By submitting a comment you agree to abide by our [Terms](#) and [Community Guidelines](#). If you find something abusive or that does not comply with our terms or guidelines please flag it as inappropriate.

[Access through your institution](#)

[Change institution](#)

[Buy or subscribe](#)

## Associated Content

### [Iron and nickel vapours are present in most comets](#)

- Dennis Bodewits
- Steven J. Bromley

Nature News & Views 19 May 2021

Advertisement

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

- Article
- [Published: 19 May 2021](#)

# Precision test of statistical dynamics with state-to-state ultracold chemistry

- [Yu Liu](#) [ORCID: orcid.org/0000-0002-5228-1997<sup>1,2,3</sup>](#) na1 nAff6,
- [Ming-Guang Hu](#) [ORCID: orcid.org/0000-0002-7728-7500<sup>1,2,3</sup>](#) na1,
- [Matthew A. Nichols](#) [ORCID: orcid.org/0000-0002-6686-0252<sup>1,2,3</sup>](#),
- [Dongzheng Yang](#) [ORCID: orcid.org/0000-0002-8567-1051<sup>4</sup>](#),
- [Daiqian Xie<sup>4</sup>](#),
- [Hua Guo](#) [ORCID: orcid.org/0000-0001-9901-053X<sup>5</sup>](#) &
- [Kang-Kuen Ni](#) [ORCID: orcid.org/0000-0002-0537-0719<sup>1,2,3</sup>](#)

[Nature](#) volume 593, pages 379–384 (2021) [Cite this article](#)

- 1010 Accesses
- 73 Altmetric
- [Metrics details](#)

## Subjects

- [Chemical physics](#)
- [Reaction kinetics and dynamics](#)

## Abstract

Chemical reactions represent a class of quantum problems that challenge both the current theoretical understanding and computational capabilities<sup>1</sup>. Reactions that occur at ultralow temperatures provide an ideal testing ground for quantum chemistry and scattering theories, because they can be experimentally studied with unprecedented control<sup>2</sup>, yet display dynamics that are highly complex<sup>3</sup>. Here we report the full product state distribution for the reaction  $2\text{KRb} \rightarrow \text{K}_2 + \text{Rb}_2$ . Ultracold preparation of the reactants allows us complete control over their initial quantum degrees of freedom, whereas state-resolved, coincident detection of both products enables the probability of scattering into each of the 57 allowed rotational state-pairs to be measured. Our results show an overall agreement with a state-counting model based on statistical theory<sup>4,5,6</sup>, but also reveal several deviating state-pairs. In particular, we observe a strong suppression of population in the state-pair closest to the exoergicity limit as a result of the long-range potential inhibiting the escape of products. The completeness of our measurements provides a benchmark for quantum dynamics calculations beyond the current state of the art.

## Access options

Subscribe to Journal

Get full journal access for 1 year

\$199.00

only \$3.90 per issue

[Subscribe](#)

All prices are NET prices.

VAT will be added later in the checkout.

Tax calculation will be finalised during checkout.

Rent or Buy article

Get time limited or full article access on ReadCube.

from \$8.99

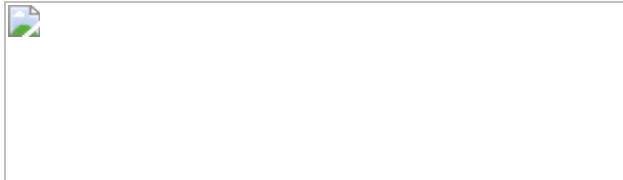
[Rent or Buy](#)

All prices are NET prices.

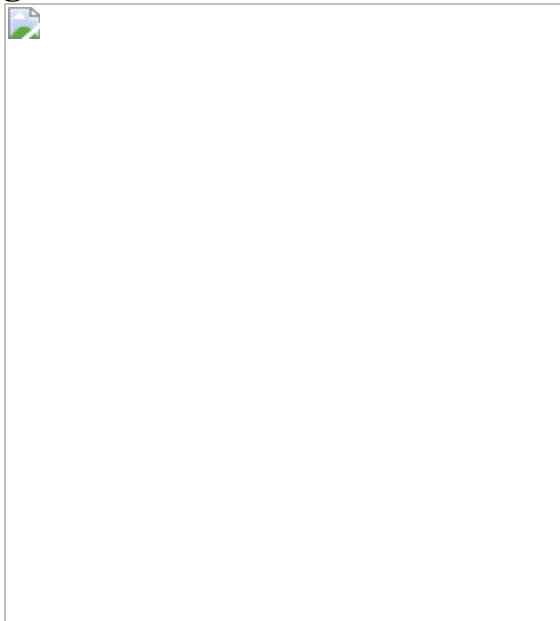
### **Additional access options:**

- [Log in](#)
- [Access through your institution](#)
- [Learn about institutional subscriptions](#)

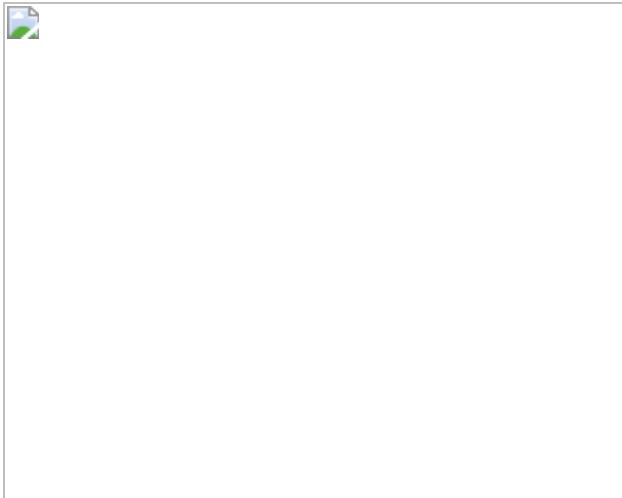
**Fig. 1: Energetics and product quantum states for ultracold reactions between KRb molecules.**



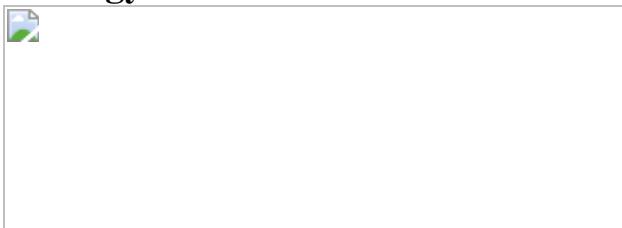
**Fig. 2: State-resolved coincidence detection of reaction products.**



**Fig. 3: Measured product state distribution and comparison to statistical theory.**



**Fig. 4: Influence of the long-range potential on product formation near the energy threshold.**



## Data availability

The data that support the findings of this study are available from the corresponding author upon reasonable request.

## Code availability

The computer codes used for theoretical calculations in this study are available from H.G. ([hguo@unm.edu](mailto:hguo@unm.edu)) upon reasonable request.

## References

1. 1.

- Li, J., Zhao, B., Xie, D. & Guo, H. Advances and new challenges to bimolecular reaction dynamics theory. *J. Phys. Chem. Lett.* **11**, 8844–8860 (2020).

[CAS](#) [Google Scholar](#)

2. 2.

Quéméner, G. & Julienne, P. S. Ultracold molecules under control! *Chem. Rev.* **112**, 4949–5011 (2012).

[Google Scholar](#)

3. 3.

Croft, J. et al. Universality and chaoticity in ultracold K + KRb chemical reactions. *Nat. Commun.* **8**, 15897 (2017).

[ADS](#) [CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

4. 4.

Light, J. C. Statistical theory of bimolecular exchange reactions. *Discuss. Faraday Soc.* **44**, 14–29 (1967).

[Google Scholar](#)

5. 5.

Nikitin, E. E. & Umanskii, S. Y. *Theory of Slow Atomic Collisions* Vol. 30 (Springer Science & Business Media, 2012).

6. 6.

Pechukas, P. in *Dynamics of Molecular Collisions* 269–322 (Springer, 1976).

7. 7.

Julienne, P. S. Ultracold molecules from ultracold atoms: a case study with the KRb molecule. *Faraday Discuss.* **142**, 361–388 (2009).

[ADS](#) [CAS](#) [Google Scholar](#)

8. 8.

Balakrishnan, N. Ultracold molecules and the dawn of cold controlled chemistry. *J. Chem. Phys.* **145**, 150901 (2016).

[ADS](#) [CAS](#) [Google Scholar](#)

9. 9.

Tarbutt, M. R. Laser cooling of molecules. *Contemp. Phys.* **59**, 356–376 (2018).

[ADS](#) [CAS](#) [Google Scholar](#)

10. 10.

Toscano, J., Lewandowski, H. & Heazlewood, B. R. Cold and controlled chemical reaction dynamics. *Phys. Chem. Chem. Phys.* **22**, 9180–9194 (2020).

[CAS](#) [Google Scholar](#)

11. 11.

Ospelkaus, S. et al. Quantum-state controlled chemical reactions of ultracold potassium-rubidium molecules. *Science* **327**, 853–857 (2010).

[ADS](#) [CAS](#) [Google Scholar](#)

12. 12.

Ni, K.-K. et al. Dipolar collisions of polar molecules in the quantum regime. *Nature* **464**, 1324 (2010).

[ADS](#) [CAS](#) [Google Scholar](#)

13. 13.

Guo, M. et al. Dipolar collisions of ultracold ground-state bosonic molecules. *Phys. Rev. X* **8**, 041044 (2018).

[CAS](#) [Google Scholar](#)

14. 14.

Yang, H. et al. Observation of magnetically tunable Feshbach resonances in ultracold  $^{23}\text{Na}^{40}\text{K} + ^{40}\text{K}$  collisions. *Science* **363**, 261–264 (2019).

[ADS](#) [CAS](#) [Google Scholar](#)

15. 15.

de Jongh, T. et al. Imaging the onset of the resonance regime in low-energy NO-He collisions. *Science* **368**, 626–630 (2020).

[ADS](#) [Google Scholar](#)

16. 16.

Rui, J. et al. Controlled state-to-state atom-exchange reaction in an ultracold atom-dimer mixture. *Nat. Phys.* **13**, 699 (2017).

[CAS](#) [Google Scholar](#)

17. 17.

Wolf, J. et al. State-to-state chemistry for three-body recombination in an ultracold rubidium gas. *Science* **358**, 921–924 (2017).

[ADS](#) [CAS](#) [Google Scholar](#)

18. 18.

Yang, X. State-to-state dynamics of elementary bimolecular reactions. *Annu. Rev. Phys. Chem.* **58**, 433–459 (2007).

[ADS](#) [CAS](#) [Google Scholar](#)

19. 19.

Clary, D. C. Theoretical studies on bimolecular reaction dynamics. *Proc. Natl Acad. Sci. USA* **105**, 12649–12653 (2008).

[ADS](#) [CAS](#) [Google Scholar](#)

20. 20.

Croft, J. F. & Bohn, J. L. Long-lived complexes and chaos in ultracold molecular collisions. *Phys. Rev. A* **89**, 012714 (2014).

[ADS](#) [Google Scholar](#)

21. 21.

Liu, Y. et al. Photo-excitation of long-lived transient intermediates in ultracold reactions. *Nat. Phys.* **16**, 1132–1136 (2020).

[ADS](#) [CAS](#) [Google Scholar](#)

22. 22.

Gregory, P. D., Blackmore, J. A., Bromley, S. L. & Cornish, S. L. Loss of ultracold  $^{87}\text{Rb}^{123}\text{Cs}$  molecules via optical excitation of long-lived two-body collision complexes. *Phys. Rev. Lett.* **124**, 163402 (2020).

[ADS](#) [CAS](#) [Google Scholar](#)

23. 23.

Bonnet, L. & Rayez, J. C. Some key factors of energy distributions in the products of complex-forming elementary reactions. *Phys. Chem. Chem. Phys.* **1**, 2383–2400 (1999).

[CAS](#) [Google Scholar](#)

24. 24.

Balucani, N. et al. Experimental and theoretical differential cross sections for the N(<sup>2</sup>D)+ H<sub>2</sub> reaction. *J. Phys. Chem. A* **110**, 817–829 (2006).

[CAS](#) [Google Scholar](#)

25. 25.

Sun, Z. et al. State-to-state dynamics of H + O<sub>2</sub> reaction, evidence for nonstatistical behavior. *J. Am. Chem. Soc.* **130**, 14962–14963 (2008).

[CAS](#) [Google Scholar](#)

26. 26.

Rivero-Santamaría, A. et al. The O(<sup>1</sup>D)+ H<sub>2</sub> ( $X^1\Sigma^+$ , v, j) → OH ( $X^2\Pi$ , v', j')+ H(<sup>2</sup>S) reaction at low collision energy: when a simple statistical description of the dynamics works. *Phys. Chem. Chem. Phys.* **13**, 8136–8139 (2011).

[Google Scholar](#)

27. 27.

González-Martínez, M. L., Dulieu, O., Larrégaray, P. & Bonnet, L. Statistical product distributions for ultracold reactions in external fields. *Phys. Rev. A* **90**, 052716 (2014).

[ADS](#) [Google Scholar](#)

28. 28.

Nesbitt, D. J. Toward state-to-state dynamics in ultracold collisions: lessons from high-resolution spectroscopy of weakly bound molecular complexes. *Chem. Rev.* **112**, 5062–5072 (2012).

[CAS](#) [Google Scholar](#)

29. 29.

Ni, K.-K. et al. A high phase-space-density gas of polar molecules. *Science* **322**, 231–235 (2008).

[ADS](#) [CAS](#) [Google Scholar](#)

30. 30.

Liu, Y., Grimes, D. D., Hu, M.-G. & Ni, K.-K. Probing ultracold chemistry using ion spectrometry. *Phys. Chem. Chem. Phys.* **22**, 4861–4874 (2020).

[CAS](#) [Google Scholar](#)

31. 31.

Hu, M.-G. et al. Direct observation of bimolecular reactions of ultracold KRb molecules. *Science* **366**, 1111–1115 (2019).

[ADS](#) [CAS](#) [Google Scholar](#)

32. 32.

Hu, M.-G. et al. Nuclear spin conservation enables state-to-state control of ultracold molecular reactions. *Nat. Chem.* <https://doi.org/10.1038/s41557-020-00610-0> (2020).

33. 33.

Byrd, J. N., Montgomery, J. A. Jr & Côté, R. Structure and thermochemistry of K<sub>2</sub>Rb, KRb<sub>2</sub>, and K<sub>2</sub>Rb<sub>2</sub>. *Phys. Rev. A* **82**, 010502 (2010).

[ADS](#) [Google Scholar](#)

34. 34.

Yang, D. et al. A global full-dimensional potential energy surface for the K<sub>2</sub>Rb<sub>2</sub> complex and its lifetime. *J. Phys. Chem. Lett.* **11**, 2605–2610 (2020).

[CAS](#) [Google Scholar](#)

35. 35.

Liu, K. Product pair correlation in bimolecular reactions. *Phys. Chem. Chem. Phys.* **9**, 17–30 (2007).

[CAS](#) [Google Scholar](#)

36. 36.

Brouard, M., O’Keeffe, P. & Vallance, C. Product state resolved dynamics of elementary reactions. *J. Phys. Chem. A* **106**, 3629–3641 (2002).

[CAS](#) [Google Scholar](#)

37. 37.

Continetti, R. E. Coincidence spectroscopy. *Annu. Rev. Phys. Chem.* **52**, 165–192 (2001).

[ADS](#) [CAS](#) [Google Scholar](#)

38. 38.

Lee, S. K. et al. Coincidence ion imaging with a fast frame camera. *Rev. Sci. Instrum.* **85**, 123303 (2014).

[ADS](#) [Google Scholar](#)

39. 39.

Wasserman, L. *All Of Statistics: A Concise Course In Statistical Inference* (Springer, 2004).

40. 40.

Meyer, E. R. & Bohn, J. L. Product-state control of bi-alkali-metal chemical reactions. *Phys. Rev. A* **82**, 042707 (2010).

[ADS](#) [Google Scholar](#)

41. 41.

Amiot, C., Vergès, J. & Fellows, C. E. The long-range potential of the  $K_2(^1\Sigma_g^+)$  ground electronic state up to 15 Å. *J. Chem. Phys.* **103**, 3350–3356 (1995).

42. 42.

Seto, J. Y., Le Roy, R. J., Verges, J. & Amiot, C. Direct potential fit analysis of the  $\psi_1(^1\Sigma_g^+)$  state of  $Rb_2$ : nothing else will do! *J. Chem. Phys.* **113**, 3067–3076 (2000).

43. 43.

Yang, D., Huang, J., Hu, X., Xie, D. & Guo, H. Statistical quantum mechanical approach to diatom–diatom capture dynamics and application to ultracold KRb + KRb reaction. *J. Chem. Phys.* **152**, 241103 (2020).

[CAS](#) [Google Scholar](#)

44. 44.

Brumer, P., Bergmann, K. & Shapiro, M. Identical collision partners in the coherent control of bimolecular reactions. *J. Chem. Phys.* **113**, 2053–2055 (2000).

[ADS](#) [CAS](#) [Google Scholar](#)

45. 45.

Kendrick, B., Hazra, J. & Balakrishnan, N. The geometric phase controls ultracold chemistry. *Nat. Commun.* **6**, 7918 (2015).

[ADS](#) [CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

46. 46.

Falke, S., Sherstov, I., Tiemann, E. & Lisdat, C. The  $\langle\{A\}^{\{1\}}\{\Sigma_g^+\}\rangle$  state of K<sub>2</sub> up to the dissociation limit. *J. Chem. Phys.* **125**, 224303 (2006).

47. 47.

Amiot, C. Laser-induced fluorescence of Rb<sub>2</sub>: the (1)  $\langle\{1\}\{\Sigma_g^+\}(X)\rangle$ , (2)  $\langle\{1\}\{\Sigma_g^+(+)\}\rangle$ , (1)  $\langle\{1\}\{\Pi_u^-(B)\}\rangle$ , (1)  $\langle\{1\}\{\Pi_g^-\}\rangle$ , and (2)  $\langle\{1\}\{\Pi_g^-(C)\}\rangle$  electronic states. *J. Chem. Phys.* **93**, 8591–8604 (1990).

48. 48.

Aldegunde, J. & Hutson, J. M. Hyperfine energy levels of alkali-metal dimers: ground-state homonuclear molecules in magnetic fields. *Phys. Rev. A* **79**, 013401 (2009).

[ADS](#) [Google Scholar](#)

49. 49.

Caldwell, C., Engelke, F. & Hage, H. High resolution spectroscopy in supersonic nozzle beams: the Rb<sub>2</sub>  $\langle\{B\}^{\{1\}}\{\Pi_u^-\}-\{X\}^{\{1\}}\{\Sigma_g^+\}\rangle$  band system. *Chem. Phys.* **54**, 21–31 (1980).

50. 50.

Amiot, C. & Vergès, J. Optical-optical double resonance and Fourier transform spectroscopy: the Rb<sub>2</sub>  $\langle\{B\}_{\{1\}}\{\Pi_u^-\}\rangle$  electronic state up to the quasibound energy levels. *Chem. Phys. Lett.* **274**, 91–98 (1997).

51. 51.

Ullrich, J. et al. Recoil-ion and electron momentum spectroscopy: reaction-microscopes. *Rep. Prog. Phys.* **66**, 1463 (2003).

[ADS](#) [CAS](#) [Google Scholar](#)

52. 52.

Vredenborg, A., Roeterdink, W. G. & Janssen, M. H. A photoelectron-photoion coincidence imaging apparatus for femtosecond time-resolved molecular dynamics with electron time-of-flight resolution of  $\sigma = 18$  ps and energy resolution  $\Delta E/E = 3.5\%$ . *Rev. Sci. Instrum.* **79**, 063108 (2008).

[ADS](#) [Google Scholar](#)

53. 53.

Eppink, A. T. & Parker, D. H. Velocity map imaging of ions and electrons using electrostatic lenses: application in photoelectron and photofragment ion imaging of molecular oxygen. *Rev. Sci. Instrum.* **68**, 3477–3484 (1997).

[ADS](#) [CAS](#) [Google Scholar](#)

54. 54.

Gericke, K.-H. Correlations between quantum state populations of coincident product pairs. *Phys. Rev. Lett.* **60**, 561 (1988).

[ADS](#) [CAS](#) [Google Scholar](#)

55. 55.

Lin, J. J., Zhou, J., Shiu, W. & Liu, K. State-specific correlation of coincident product pairs in the F + CD<sub>4</sub> reaction. *Science* **300**, 966–969 (2003).

[ADS](#) [CAS](#) [Google Scholar](#)

56. 56.

Dixit, S., Levin, D. & McKoy, B. Resonant enhanced multiphoton ionization studies in atomic oxygen. *Phys. Rev. A* **37**, 4220 (1988).

[ADS](#) [CAS](#) [Google Scholar](#)

57. 57.

Johnson, B. An improved log derivative method for inelastic scattering. *J. Comput. Phys.* **13**, 445–449 (1973).

[ADS](#) [Google Scholar](#)

58. 58.

Manolopoulos, D. An improved log derivative method for inelastic scattering. *J. Chem. Phys.* **85**, 6425–6429 (1986).

[ADS](#) [CAS](#) [Google Scholar](#)

59. 59.

Johnson, B. A generalized JWKB approximation for multichannel scattering. *Chem. Phys.* **2**, 381–399 (1973).

[Google Scholar](#)

[Download references](#)

## Acknowledgements

We thank L. Zhu for experimental assistance; T. Rosenband, G. Quéméner, W. Cairncross, E. Heller and M. Soley for discussions; T. Karman for providing the code for state-counting; J. Huang for ab initio calculations; and L. Liu for a critical reading of the manuscript. This work

is supported by the DOE Young Investigator Program (DE-SC0019020) and the David and Lucile Packard Foundation. M.A.N. is supported by the Arnold O. Beckman Postdoctoral Fellowship in Chemical Instrumentation. D.Y. and D.X. acknowledge support from the National Natural Science Foundation of China (grant numbers 21733006). H.G. thanks the Army Research Office (W911NF-19-1-0283) for funding.

## Author information

### Author notes

1. Yu Liu

Present address: Time and Frequency Division, National Institute of Standards and Technology, Boulder, CO, USA

2. These authors contributed equally: Yu Liu, Ming-Guang Hu

### Affiliations

1. Department of Chemistry and Chemical Biology, Harvard University, Cambridge, MA, USA

Yu Liu, Ming-Guang Hu, Matthew A. Nichols & Kang-Kuen Ni

2. Department of Physics, Harvard University, Cambridge, MA, USA

Yu Liu, Ming-Guang Hu, Matthew A. Nichols & Kang-Kuen Ni

3. Harvard-MIT Center for Ultracold Atoms, Cambridge, MA, USA

Yu Liu, Ming-Guang Hu, Matthew A. Nichols & Kang-Kuen Ni

4. Institute of Theoretical and Computational Chemistry, School of Chemistry and Chemical Engineering, Nanjing University, Nanjing, China

Dongzheng Yang & Daiqian Xie

5. Department of Chemistry and Chemical Biology, University of New Mexico, Albuquerque, NM, USA

Hua Guo

Authors

1. Yu Liu

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

2. Ming-Guang Hu

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

3. Matthew A. Nichols

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

4. Dongzheng Yang

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

5. Daiqian Xie

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

6. Hua Guo

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

7. Kang-Kuen Ni

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

## Contributions

The experimental work and data analysis were carried out by Y.L., M.-G.H., M.A.N. and K.-K.N. Theoretical calculations were performed by D.Y., D.X. and H.G. All authors contributed to interpreting the results and writing the manuscript.

## Corresponding authors

Correspondence to [Yu Liu](#) or [Kang-Kuen Ni](#).

## Ethics declarations

## Competing interests

The authors declare no competing interests.

## Additional information

**Peer review information** *Nature* thanks Simon Cornish, Nandini Mukherjee and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

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

## Extended data figures and tables

### [Extended Data Fig. 1 Timing diagram for product ionization.](#)

The relative timing between the ODT, REMPI, and cleanup pulses during the state-selective ionization of reaction products. The inset shows a close-up view of a single modulation period. Unperturbed reactions occur during

the dark phase of the period, while reactions influenced by the (1,064 nm) ODT light occur during the bright phase. The numbers in parentheses indicate pulse durations.

### Extended Data Fig. 2 Modelling the geometric factor for product sampling.

The plot shows the fraction of product pairs that remain within the REMPI beam volume at the time of ionization as a function of the velocity of the K<sub>2</sub> product within the pair. Blue and red dashed lines indicate the maximum velocities of the K<sub>2</sub> and Rb<sub>2</sub> products, respectively. The inset shows the ionization geometry.

### Extended Data Fig. 3 Modelling the Doppler factor for product sampling.

**a**, Normalized optical intensity profiles of the REMPI laser pulses. The red trace corresponds to the 648 nm or 674 nm lights, and is proportional to  $\langle \varOmega_{01} \rangle$ . The green trace corresponds to the 532 nm light, and is proportional to  $\langle \varGamma_{\text{ion}} \rangle$ . **b**, The Doppler factor  $F_{\text{Doppler}}(v)$  versus the velocity of the K<sub>2</sub> product. The lower and upper bounds correspond to  $\langle \varGamma_{i0} \rangle / (2\pi)$  at peak values of 6 MHz and 14 MHz, respectively.

### Extended Data Fig. 4 Distribution of product translational energy.

The measured (blue circle) and predicted (red diamond) scattering probabilities for all allowed state-pairs are plotted versus their translational energies ( $T$ ). The two sets of points are offset horizontally by 0.014 cm<sup>-1</sup> for clarity. To aid in the identification of systematic deviations, we multiply each scattering probability by a normalized Gaussian function with a  $1\sigma$  width of 0.25 cm<sup>-1</sup>, and sum them up to construct broadened distributions

as shown by the blue and red curves. These curves are scaled by a factor of 0.2 for convenience.

**Extended Data Table 1 Molecular dissociation energies of  $^{40}\text{K}^{87}\text{Rb}$ ,  $^{40}\text{K}_2$  and  $^{87}\text{Rb}_2$**

[Full size table](#)

**Extended Data Table 2 Rotational and centrifugal distortion constants for  $^{40}\text{K}_2$  and  $^{87}\text{Rb}_2$**

[Full size table](#)

**Extended Data Table 3 The internal energy ( $U$ ) and degeneracy ( $\{\mathcal{D}\}$ ) for all measured product state-pairs**

[Full size table](#)

**Extended Data Table 4 Various measured quantities that are used towards calculating the scattering probability into each product state-pair**

[Full size table](#)

## Rights and permissions

[Reprints and Permissions](#)

## About this article



Check for  
updates

## Cite this article

Liu, Y., Hu, MG., Nichols, M.A. *et al.* Precision test of statistical dynamics with state-to-state ultracold chemistry. *Nature* **593**, 379–384 (2021).  
<https://doi.org/10.1038/s41586-021-03459-6>

[Download citation](#)

- Received: 12 January 2021
- Accepted: 16 March 2021
- Published: 19 May 2021
- Issue Date: 20 May 2021
- DOI: <https://doi.org/10.1038/s41586-021-03459-6>

## Comments

By submitting a comment you agree to abide by our [Terms](#) and [Community Guidelines](#). If you find something abusive or that does not comply with our terms or guidelines please flag it as inappropriate.

[Access through your institution](#)

[Change institution](#)

[Buy or subscribe](#)

## Associated Content

### [Ultracold chemical reactions reveal the quantum mechanism of product formation](#)

- Nandini Mukherjee

Nature News & Views 19 May 2021

Advertisement

---

This article was downloaded by **calibre** from <https://www.nature.com/articles/s41586-021-03459-6>

- Article
- [Published: 19 May 2021](#)

# Tunable non-integer high-harmonic generation in a topological insulator

- [C. P. Schmid](#)<sup>1</sup>,
- [L. Weigl](#)<sup>1</sup>,
- [P. Grössing](#)<sup>2</sup>,
- [V. Junk](#) [ORCID: orcid.org/0000-0003-1675-8548](#)<sup>2</sup>,
- [C. Gorini](#)<sup>2</sup> [nAff7](#),
- [S. Schlauderer](#)<sup>1</sup>,
- [S. Ito](#)<sup>3</sup>,
- [M. Meierhofer](#)<sup>1</sup>,
- [N. Hofmann](#)<sup>1</sup>,
- [D. Afanasiev](#) [ORCID: orcid.org/0000-0001-6726-3479](#)<sup>1</sup>,
- [J. Crewse](#)<sup>2</sup>,
- [K. A. Kokh](#) [ORCID: orcid.org/0000-0003-1967-9642](#)<sup>4,5</sup>,
- [O. E. Tereshchenko](#) [ORCID: orcid.org/0000-0002-6157-5874](#)<sup>5,6</sup>,
- [J. Gütde](#)<sup>3</sup>,
- [F. Evers](#)<sup>2</sup>,
- [J. Wilhelm](#) [ORCID: orcid.org/0000-0001-8678-8246](#)<sup>2</sup>,
- [K. Richter](#) <sup>2</sup>,
- [U. Höfer](#)<sup>3</sup> &
- [R. Huber](#) [ORCID: orcid.org/0000-0001-6617-9283](#)<sup>1</sup>

[Nature](#) volume **593**, pages 385–390 (2021) [Cite this article](#)

- 1465 Accesses

- 46 Altmetric
- [Metrics details](#)

## Subjects

- [High-harmonic generation](#)
- [Topological insulators](#)

## Abstract

When intense lightwaves accelerate electrons through a solid, the emerging high-order harmonic (HH) radiation offers key insights into the material<sup>1,2,3,4,5,6,7,8,9,10,11</sup>. Sub-optical-cycle dynamics—such as dynamical Bloch oscillations<sup>2,3,4,5</sup>, quasiparticle collisions<sup>6,12</sup>, valley pseudospin switching<sup>13</sup> and heating of Dirac gases<sup>10</sup>—leave fingerprints in the HH spectra of conventional solids. Topologically non-trivial matter<sup>14,15</sup> with invariants that are robust against imperfections has been predicted to support unconventional HH generation<sup>16,17,18,19,20</sup>. Here we experimentally demonstrate HH generation in a three-dimensional topological insulator—bismuth telluride. The frequency of the terahertz driving field sharply discriminates between HH generation from the bulk and from the topological surface, where the unique combination of long scattering times owing to spin–momentum locking<sup>17</sup> and the quasi-relativistic dispersion enables unusually efficient HH generation. Intriguingly, all observed orders can be continuously shifted to arbitrary non-integer multiples of the driving frequency by varying the carrier-envelope phase of the driving field—in line with quantum theory. The anomalous Berry curvature warranted by the non-trivial topology enforces meandering ballistic trajectories of the Dirac fermions, causing a hallmark polarization pattern of the HH emission. Our study provides a platform to explore topology and relativistic quantum physics in strong-field control, and could lead to non-dissipative topological electronics at infrared frequencies.

## Access options

[Subscribe to Journal](#)

Get full journal access for 1 year

\$199.00

only \$3.90 per issue

[Subscribe](#)

All prices are NET prices.

VAT will be added later in the checkout.

Tax calculation will be finalised during checkout.

Rent or Buy article

Get time limited or full article access on ReadCube.

from \$8.99

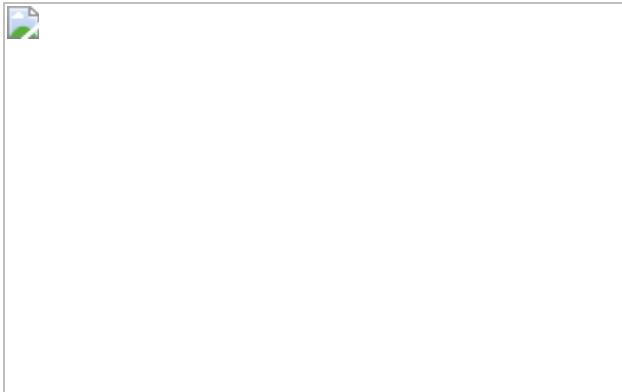
[Rent or Buy](#)

All prices are NET prices.

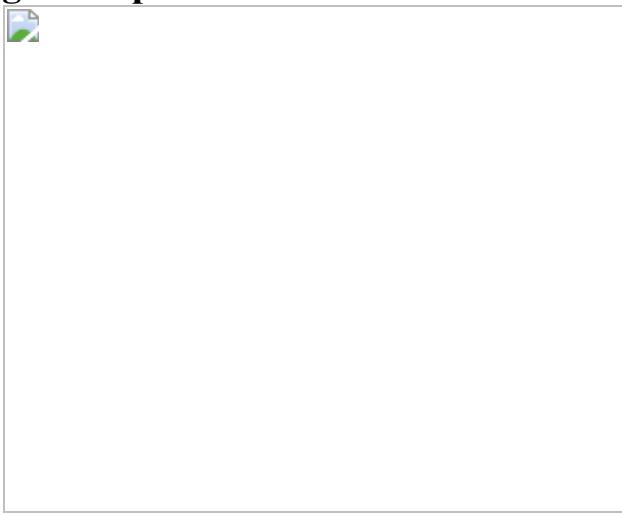
### **Additional access options:**

- [Log in](#)
- [Access through your institution](#)
- [Learn about institutional subscriptions](#)

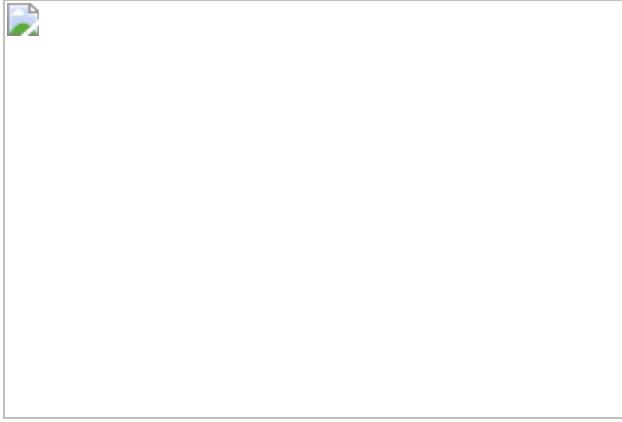
**Fig. 1: HH emission from a TI.**



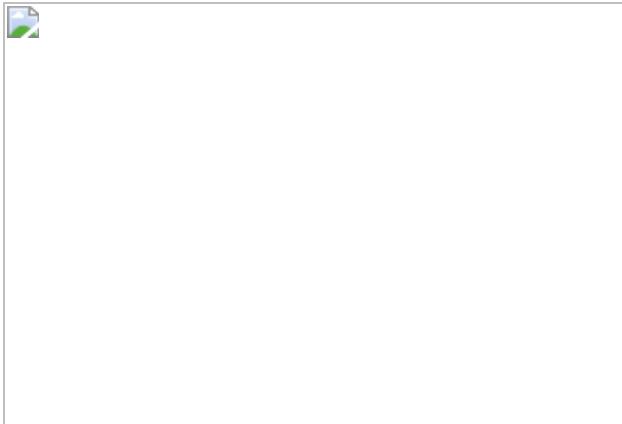
**Fig. 2: Dependence of the HH emission on the CEP of the driving field.**



**Fig. 3: Microscopic origin of high-efficiency HHG from the TSS.**



**Fig. 4: Tracing geometric phase effects in HH radiation from the TSS.**



## Data availability

The data supporting the findings of this study are available from the corresponding authors upon request. [Source data](#) are provided with this paper.

## Code availability

The in-house program package CUED that was used to solve the SBE is freely available from GitHub (<https://github.com/ccmt-regensburg/CUED>).

## References

1. 1.

Chin, A. H., Calderón, O. G. & Kono, J. Extreme midinfrared nonlinear optics in semiconductors. *Phys. Rev. Lett.* **86**, 3292–3295 (2001).

[ADS](#) [CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

2. 2.

Ghimire, S. et al. Observation of high-order harmonic generation in a bulk crystal. *Nat. Phys.* **7**, 138–141 (2011).

[CAS](#) [Google Scholar](#)

3. 3.

Schubert, O. et al. Sub-cycle control of terahertz high-harmonic generation by dynamical Bloch oscillations. *Nat. Photon.* **8**, 119–123 (2014).

[ADS](#) [CAS](#) [Google Scholar](#)

4. 4.

Hohenleutner, M. et al. Real-time observation of interfering crystal electrons in high-harmonic generation. *Nature* **523**, 572–575 (2015).

[ADS](#) [CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

5. 5.

Luu, T. T. et al. Extreme ultraviolet high-harmonic spectroscopy of solids. *Nature* **521**, 498–502 (2015).

[ADS](#) [CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

6. 6.

Vampa, G. et al. Linking high harmonics from gases and solids. *Nature* **522**, 462–464 (2015); corrigendum **542**, 260 (2017).

[ADS](#) [CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

7. 7.

Garg, M. et al. Multi-petahertz electronic metrology. *Nature* **538**, 359–363 (2016).

[ADS](#) [CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

8. 8.

Yoshikawa, N., Tamaya, T. & Tanaka, K. High-harmonic generation in graphene enhanced by elliptically polarized light excitation. *Science* **356**, 736–738 (2017).

[ADS](#) [MathSciNet](#) [CAS](#) [PubMed](#) [PubMed Central](#) [MATH](#) [Google Scholar](#)

9. 9.

Sivis, M. et al. Tailored semiconductors for high-harmonic optoelectronics. *Science* **357**, 303–306 (2017).

[ADS](#) [CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

10. 10.

Hafez, H. A. et al. Extremely efficient terahertz high-harmonic generation in graphene by hot Dirac fermions. *Nature* **561**, 507–511 (2018).

[ADS](#) [CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

11. 11.

Floss, I. et al. Ab initio multiscale simulation of high-order harmonic generation in solids. *Phys. Rev A* **97**, 011401(R) (2018).

[ADS](#) [Google Scholar](#)

12. 12.

Langer, F. et al. Lightwave-driven quasiparticle collisions on a subcycle timescale. *Nature* **533**, 225–229 (2016).

[ADS](#) [CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

13. 13.

Langer, F. et al. Lightwave valleytronics in a monolayer of tungsten diselenide. *Nature* **557**, 76–80 (2018).

[ADS](#) [CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

14. 14.

Chen, Y. L. et al. Experimental realization of a three-dimensional topological insulator,  $\text{Bi}_2\text{Te}_3$ . *Science* **325**, 178–181 (2009).

[ADS](#) [CAS](#) [Google Scholar](#)

15. 15.

Giorgianni, F. et al. Strong nonlinear terahertz response induced by Dirac surface states in  $\text{Bi}_2\text{Se}_3$  topological insulator. *Nat. Commun.* **7**, 11421 (2016).

[ADS](#) [CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

16. 16.

Bauer, D. & Hansen, K. K. High-harmonic generation in solids with and without topological edge states. *Phys. Rev. Lett.* **120**, 177401 (2018).

[ADS](#) [CAS](#) [Google Scholar](#)

17. 17.

Reimann, J. et al. Subcycle observation of lightwave-driven Dirac currents in a topological surface band. *Nature* **562**, 396–400 (2018).

[ADS](#) [CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

18. 18.

Silva, R. E. F., Jiménez-Galán, A., Amorim, B., Smirnova, O. & Ivanov, M. Topological strong field physics on sub-laser cycle

timescale. *Nat. Photon.* **13**, 849–854 (2019).

[ADS](#) [CAS](#) [Google Scholar](#)

19. 19.

Baykusheva, D. et al. Strong-field physics in three-dimensional topological insulators. *Phys. Rev. A* **103**, 023101 (2021).

[ADS](#) [CAS](#) [Google Scholar](#)

20. 20.

Wilhelm, J. et al. Semiconductor-Bloch formalism: derivation and application to high-harmonic generation from Dirac fermions. *Phys. Rev. B* **103**, 125419 (2021).

[ADS](#) [CAS](#) [Google Scholar](#)

21. 21.

Higuchi, T., Heide, C., Ullmann, K., Weber, H. B. & Hommelhoff, P. Light-field-driven currents in graphene. *Nature* **550**, 224–228 (2017).

[ADS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

22. 22.

McIver, J. W. et al. Light-induced anomalous Hall effect in graphene. *Nat. Phys.* **16**, 38–41 (2020).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

23. 23.

Kuroda, K., Reimann, J., Gündde, J. & Höfer, U. Generation of transient photocurrents in the topological surface state of  $\text{Sb}_2\text{Te}_3$  by direct optical excitation with midinfrared pulses. *Phys. Rev. Lett.* **116**, 076801 (2016).

[ADS](#) [CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

24. 24.

Wu, L. et al. Quantized Faraday and Kerr rotation and axion electrodynamics of a 3D topological insulator. *Science* **354**, 1124–1127 (2016).

[ADS](#) [MathSciNet](#) [CAS](#) [PubMed](#) [PubMed Central](#) [MATH](#) [Google Scholar](#)

25. 25.

Mahmood, F. et al. Selective scattering between Floquet–Bloch and Volkov states in a topological insulator. *Nat. Phys.* **12**, 306–310 (2016).

[CAS](#) [Google Scholar](#)

26. 26.

Braun, L. et al. Ultrafast photocurrents at the surface of the three-dimensional topological insulator Bi<sub>2</sub>Se<sub>3</sub>. *Nat. Commun.* **7**, 13259 (2016).

[ADS](#) [CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

27. 27.

Berry, M. V. Quantal phase factors accompanying adiabatic changes. *Proc. R. Soc. Lond. A* **392**, 45–57 (1984).

[ADS](#) [MathSciNet](#) [MATH](#) [Google Scholar](#)

28. 28.

Luu, T. T. & Wörner, H. J. Measurement of the Berry curvature of solids using high-harmonic spectroscopy. *Nat. Commun.* **9**, 916 (2018).

[ADS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

29. 29.

Banks, H. B. et al. Dynamical birefringence: electron–hole recollisions as probes of Berry curvature. *Phys. Rev. X* **7**, 041042 (2017).

[Google Scholar](#)

30. 30.

Cheng, B. et al. Efficient terahertz harmonic generation with coherent acceleration of electrons in the Dirac semimetal Cd<sub>3</sub>As<sub>2</sub>. *Phys. Rev. Lett.* **124**, 117402 (2020).

[ADS](#) [CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

31. 31.

Sell, A., Leitenstorfer, A. & Huber, R. Phase-locked generation and field-resolved detection of widely tunable terahertz pulses with amplitudes exceeding 100 MV/cm. *Opt. Lett.* **33**, 2767–2769 (2008).

[ADS](#) [CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

32. 32.

Michiardi, M. et al. Bulk band structure of Bi<sub>2</sub>Te<sub>3</sub>. *Phys. Rev. B* **90**, 075105 (2014).

[ADS](#) [CAS](#) [Google Scholar](#)

33. 33.

Austin, I. G. The optical properties of bismuth telluride. *Proc. Phys. Soc.* **72**, 545–552 (1958).

[ADS](#) [CAS](#) [Google Scholar](#)

34. 34.

Kokh, K. A. et al. Melt growth of bulk Bi<sub>2</sub>Te<sub>3</sub> crystals with a natural p–n junction. *CrystEngComm* **16**, 581–584 (2014).

[CAS](#) [Google Scholar](#)

35. 35.

Liu, C.-X. et al. Model Hamiltonian for topological insulators. *Phys. Rev. B* **82**, 045122 (2010).

[ADS](#) [Google Scholar](#)

36. 36.

Xiao, D., Chang, M.-C. & Niu, Q. Berry phase effects on electronic properties. *Rev. Mod. Phys.* **82**, 1959–2007 (2010).

[ADS](#) [MathSciNet](#) [CAS](#) [MATH](#) [Google Scholar](#)

37. 37.

Gradhand, M. et al. First-principle calculations of the Berry curvature of Bloch states for charge and spin transport of electrons. *J. Phys. Condens. Matter* **24**, 213202 (2012).

[ADS](#) [CAS](#) [Google Scholar](#)

38. 38.

Kane, E. O. Zener tunneling in semiconductors. *J. Phys. Chem. Solids* **12**, 181–188 (1960).

[ADS](#) [CAS](#) [Google Scholar](#)

39. 39.

Lange, C. et al. Extremely nonperturbative nonlinearities in GaAs driven by atomically strong terahertz fields in gold metamaterials. *Phys. Rev. Lett.* **113**, 227401 (2014).

[ADS](#) [CAS](#) [Google Scholar](#)

40. 40.

Junginger, F. et al. Nonperturbative interband response of a bulk InSb semiconductor driven off resonantly by terahertz electromagnetic few-cycle pulses. *Phys. Rev. Lett.* **109**, 147403 (2012).

[ADS](#) [CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

41. 41.

Kira, M. & Koch, S. W. *Semiconductor Quantum Optics* (Cambridge Univ. Press, 2012).

42. 42.

Mikhailov, S. A. Non-linear electromagnetic response of graphene. *Europhys. Lett.* **79**, 27002 (2007).

[ADS](#) [Google Scholar](#)

43. 43.

Huard, S. *Polarization of Light* (Wiley, 1997).

44. 44.

Junk, V., Reck, P., Gorini, C. & Richter, K. Floquet oscillations in periodically driven Dirac systems. *Phys. Rev. B* **101**, 134302 (2020).

[ADS](#) [CAS](#) [Google Scholar](#)

45. 45.

Krückl, V. *Wave Packets in Mesoscopic Systems: From Time-dependent Dynamics to Transport Phenomena in Graphene and Topological Insulators*. PhD thesis, Univ. Regensburg (2013).

[Download references](#)

## Acknowledgements

We thank P. Merkl, J. Freudenstein, C. Lange, D. E. Kim, M. Nitsch and I. Floss for helpful discussions. The work in Regensburg has been supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) through Project ID 422 314695032-SFB 1277 (Subprojects A03, A05 and A07) as well as project HU1598/8. Work in Marburg has been supported by the Deutsche Forschungsgemeinschaft (DFG) through Project ID 223848855-SFB 1083 and grant number GU 495/2. O.E.T. and K.A.K. have been supported by the Russian Science Foundation (project number 17-12-01047) and the state assignment of IGM SB RAS and ISP SB RAS. The work of J.C. was supported by the NSF (National Science Foundation) DMR-1828489.

## Author information

### Author notes

#### 1. C. Gorini

Present address: Université Paris-Saclay, CEA, CNRS, SPEC, Gif-sur-Yvette, France

## Affiliations

#### 1. Institute of Experimental and Applied Physics, University of Regensburg, Regensburg, Germany

C. P. Schmid, L. Weigl, S. Schlauderer, M. Meierhofer, N. Hofmann, D. Afanasiev & R. Huber

2. Institute of Theoretical Physics, University of Regensburg,  
Regensburg, Germany

P. Grössing, V. Junk, C. Gorini, J. Crewse, F. Evers, J. Wilhelm & K. Richter

3. Department of Physics, Philipps-University of Marburg, Marburg,  
Germany

S. Ito, J. Gütde & U. Höfer

4. V.S. Sobolev Institute of Geology and Mineralogy SB RAS,  
Novosibirsk, Russia

K. A. Kokh

5. Novosibirsk State University, Novosibirsk, Russia

K. A. Kokh & O. E. Tereshchenko

6. A.V. Rzhanov Institute of Semiconductor Physics SB RAS,  
Novosibirsk, Russia

O. E. Tereshchenko

## Authors

1. C. P. Schmid

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

2. L. Weigl

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

3. P. Grössing

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

4. V. Junk

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

5. C. Gorini

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

6. S. Schlauderer

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

7. S. Ito

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

8. M. Meierhofer

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

9. N. Hofmann

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

10. D. Afanasiev

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

11. J. Crewse

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

12. K. A. Kokh

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

13. O. E. Tereshchenko

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

14. J. Gündde

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

15. F. Evers

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

16. J. Wilhelm

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

17. K. Richter

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

18. U. Höfer

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

19. R. Huber

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

## Contributions

F.E., J.W., K.R., U.H. and R.H. conceived the study. K.A.K. and O.E.T. provided the high-quality Bi<sub>2</sub>Te<sub>3</sub> samples and performed transport measurements. C.P.S., L.W., S.S., S.I., M.M., N.H., D.A. and J.G. set up and carried out the optical experiments and characterized the sample orientation. C.P.S., L.W., P.G., J.C., F.E. and J.W. developed and carried out the semiconductor Bloch equations simulations. V.J., C.G. and K.R. set up and carried out the quantum mechanical wave-packet simulation. C.P.S., L.W. and P.G. implemented the semiclassical calculations. All authors analysed the data and discussed the results. C.P.S. and R.H. wrote the manuscript with contributions from all authors.

## Corresponding authors

Correspondence to [J. Wilhelm](#) or [K. Richter](#) or [R. Huber](#).

## Ethics declarations

### Competing interests

The authors declare no competing interests.

## Additional information

**Peer review information** *Nature* thanks Olga Smirnova, Ryusuke Matsunaga and Alexander Kemper for their contribution to the peer review of this work.

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

## Extended data figures and tables

## Extended Data Fig. 1 Crystallographic orientation of the Bi<sub>2</sub>Te<sub>3</sub> sample.

**a**, Low-energy electron diffraction of the Bi<sub>2</sub>Te<sub>3</sub> sample measured with an electron energy of 78 eV. The white spots mark the reciprocal lattice vectors. **b**, Schematic of the reciprocal lattice vectors (blue) overlaid with the surface Brillouin zone (red) and the corresponding high-symmetry directions.

## Extended Data Fig. 2 Carrier-injection into bulk states and comparison of bulk- and surface-state HHG.

**a**, Calculated carrier density injected, within one THz half cycle, into the bulk conduction band by Zener tunnelling, as a function of the peak THz electric field. Red horizontal line: carrier density,  $n_{\text{bulk}}$ , injected for  $E_{\text{THz}} = 3 \text{ MV cm}^{-1}$ . **b**, HH spectra,  $I_{\text{HH}}$ , calculated by Boltzmann equations for massive electrons in the bulk crystal described by a cosine-shaped band structure (blue curve), and by Boltzmann equation for Dirac electrons in the surface state of a topological insulator. Inset: corresponding band structures for bulk (blue curve) and surface state electrons (red curve). [Source Data](#)

## Extended Data Fig. 3 Comparison of the intensity of bulk and surface HHG.

**a**, HH intensity,  $I_{\text{HH}}$ , for two select driving frequencies. Above-bandgap excitation at  $\nu_{\text{THz}} = 37 \text{ THz}$  (red curve) allows for strong bulk contributions. For  $\nu_{\text{THz}} = 28 \text{ THz}$  (blue curve) resonant interband transitions in the bulk are not possible and the peak electric THz field (about 3 MV cm<sup>-1</sup>) is too low for efficient non-resonant interband excitation. Therefore, the bulk contribution drops by orders of magnitude and the observed spectrum is dominated by HHG from the metallic TSS. In a direct comparison of the two spectra, this contribution is reduced with respect to the above-bandgap bulk HH intensity by only one order of magnitude. **b**, A direct comparison of the same spectra  $I_{\text{HH}}$  as a function of the harmonic order,  $n$ , instead of the harmonic frequency, even reveals a slight enhancement of HHG in the TSS

( $v_{\text{THz}} = 28 \text{ THz}$ ) with respect to the above-bandgap bulk HHG ( $v_{\text{THz}} = 37 \text{ THz}$ ). Considering the low effective thickness of the TSS of about 1 nm compared with the optical penetration depth of about 30 nm to 100 nm over which bulk HHG is collected, this comparison attests to the strong nonlinearity of Dirac electrons. [Source Data](#)

### Extended Data Fig. 4 CEP dependence for HHG in the TSS.

HH intensity,  $I_{\text{HH}}$ , generated in the TSS as a function of the CEP,  $\varphi_{\text{CEP}}$ , along the black dotted line in the inset. The intensity of the emitted HHs monotonically increases with increasing  $\varphi_{\text{CEP}}$ . [Source Data](#)

### Extended Data Fig. 5 SBE simulation of CEP-dependent HHG from TSS without interband transitions.

Numerical simulation of  $I_{\text{HH}}$  from the TSS with the SBEs, as in Fig. 2e, but deactivated interband transitions. This calculation is equivalent to the semiclassical solution using the Boltzmann equation, which accounts only for intraband dynamics. The results reproduce both the CEP dependence observed in the experiment of Fig. 2b and the full SBE results of Fig. 2e. [Source Data](#)

### Extended Data Fig. 6 Momentum-space origin of HHG in the TSS.

**a**, Calculated HH spectra,  $I_{\text{HH}}$ , (black curve) for two test charges placed at the wave vector  $k_y = \pm 0.001 \text{ \AA}^{-1}$  ( $k_x = 0$ ), as obtained from a semiclassical solution of the equations of motion ( $v_{\text{THz}} = 25 \text{ THz}$ ,  $E_{\text{THz}} = 0.1 \text{ MV cm}^{-1}$ ). **b**, HH intensity (colour scale) of order  $n = 15$  (see arrow and red dotted area in **a**) as a function of the starting point ( $k_x$ ,  $k_y$ ) of the test charges in momentum space. [Source Data](#)

### Extended Data Fig. 7 Quantum mechanical wave-packet motion in the TSS.

**a**, Top: normalized vector potential,  $A_{\text{THz}}$ , of the driving multi-THz waveform (frequency  $\nu_{\text{THz}} = 25 \text{ THz}$ ; peak electric field  $E_{\text{THz}} = 1 \text{ MV cm}^{-1}$ ). Dashed lines highlight the zero crossings of the vector potential and the momentum space trajectories. Bottom: group velocity components of the electrons in the TSS parallel ( $v_x$ , blue) and perpendicular ( $v_y$ , red) to the THz driving field calculated by solving the full time-dependent Schrödinger equation. Both components reverse sign during zero crossings of the momentum space trajectories. **b**, Real space trajectory of lightwave-driven Dirac electrons calculated by the velocities in **a**. [Source Data](#)

### [Extended Data Fig. 8 Extended analysis of the polarimetry measurements.](#)

Extracted orientation angle,  $\alpha$ , ellipticity angle,  $\gamma$ , and degree of polarization,  $\sigma$ , as a function of the harmonic order,  $n$ . Although  $\alpha$  shows an alternating behaviour for even and odd orders, the ellipticity remains relatively small for all orders. The degree of polarization,  $\sigma$ , decreases with increasing order, but still remains sufficiently high to guarantee a reliable extraction of  $\alpha$  and  $\gamma$ . [Source Data](#)

### [Extended Data Fig. 9 Zitterbewegung in a topological surface state.](#)

Left: three-dimensional scheme of the Dirac-like electron dispersion of the TSS. The blue arrow highlights the quantum interference of different branches of the Dirac system. Right: high-frequency oscillations (blue waveform) indicative of Zitterbewegung depend on the energy separation of the interfering states residing at different energy branches in our quantum mechanical calculations. The black waveform represents the driving THz field  $E_{\text{THz}}$ . [Source Data](#)

## Source data

### [Source Data Fig. 1](#)

[Source Data Fig. 2](#)

[Source Data Fig. 3](#)

[Source Data Fig. 4](#)

[Source Data Extended Data Fig. 2](#)

[Source Data Extended Data Fig. 3](#)

[Source Data Extended Data Fig. 4](#)

[Source Data Extended Data Fig. 5](#)

[Source Data Extended Data Fig. 6](#)

[Source Data Extended Data Fig. 7](#)

[Source Data Extended Data Fig. 8](#)

[Source Data Extended Data Fig. 9](#)

## Rights and permissions

[Reprints and Permissions](#)

## About this article



Check for  
updates

[\*\*Cite this article\*\*](#)

Schmid, C.P., Weigl, L., Grössing, P. *et al.* Tunable non-integer high-harmonic generation in a topological insulator. *Nature* **593**, 385–390 (2021). <https://doi.org/10.1038/s41586-021-03466-7>

## [Download citation](#)

- Received: 28 August 2020
- Accepted: 17 March 2021
- Published: 19 May 2021
- Issue Date: 20 May 2021
- DOI: <https://doi.org/10.1038/s41586-021-03466-7>

## Comments

By submitting a comment you agree to abide by our [Terms](#) and [Community Guidelines](#). If you find something abusive or that does not comply with our terms or guidelines please flag it as inappropriate.

[Access through your institution](#)

[Change institution](#)

[Buy or subscribe](#)

Advertisement

---

This article was downloaded by **calibre** from <https://www.nature.com/articles/s41586-021-03466-7>

Discovery, characterization and engineering of ligases for amide synthesis

[Download PDF](#)

- Article
- [Published: 19 May 2021](#)

# Discovery, characterization and engineering of ligases for amide synthesis

- [Michael Winn](#)<sup>1</sup>,
- [Michael Rowlinson](#)<sup>1</sup>,
- [Fanghua Wang](#)<sup>1</sup> nAff2,
- [Luis Bering](#)<sup>1</sup>,
- [Daniel Francis](#)<sup>1</sup>,
- [Colin Levy](#)<sup>1</sup> &
- [Jason Micklefield](#)   [ORCID: orcid.org/0000-0001-8951-4873](#)<sup>1</sup>

[Nature](#) volume 593, pages 391–398 (2021) [Cite this article](#)

- 5149 Accesses
- 93 Altmetric
- [Metrics details](#)

## Subjects

- [Biocatalysis](#)
- [Biosynthesis](#)
- [Jasmonic acid](#)
- [Ligases](#)
- [Sustainability](#)

## Abstract

Coronatine and related bacterial phytotoxins are mimics of the hormone jasmonyl-l-isoleucine (JA-Ile), which mediates physiologically important plant signalling pathways<sup>1,2,3,4</sup>. Coronatine-like phytotoxins disrupt these essential pathways and have potential in the development of safer, more selective herbicides. Although the biosynthesis of coronatine has been investigated previously, the nature of the enzyme that catalyses the crucial coupling of coronafacic acid to amino acids remains unknown<sup>1,2</sup>. Here we characterize a family of enzymes, coronafacic acid ligases (CfaLs), and resolve their structures. We found that CfaL can also produce JA-Ile, despite low similarity with the Jar1 enzyme that is responsible for ligation of JA and l-Ile in plants<sup>5</sup>. This suggests that Jar1 and CfaL evolved independently to catalyse similar reactions—Jar1 producing a compound essential for plant development<sup>4,5</sup>, and the bacterial ligases producing analogues toxic to plants. We further demonstrate how CfaL enzymes can be used to synthesize a diverse array of amides, obviating the need for protecting groups. Highly selective kinetic resolutions of racemic donor or acceptor substrates were achieved, affording homochiral products. We also used structure-guided mutagenesis to engineer improved CfaL variants. Together, these results show that CfaLs can deliver a wide range of amides for agrochemical, pharmaceutical and other applications.

[Download PDF](#)

## Main

Coronatine **3** (COR) is an important phytotoxin produced by bacterial plant pathogens; it is composed of the polyketide coronafacic acid **1** (CFA), conjugated via an amide bond to coronamic acid **2** (CMA), an unusual cyclopropyl amino acid (Fig. **1a**)<sup>1,2</sup>. COR is a structural mimic of JA-Ile (**6**), a ubiquitous plant hormone that is essential for plant development and defence<sup>3</sup>. The biologically active stereoisomer is (3*R*,7*S*)-JA-Ile, but the C7 stereocentre rapidly epimerizes at physiological pH to the more stable but inactive *trans* (3*R*,7*R*) diastereoisomer, which modulates its activity<sup>4</sup>. In contrast, COR is configurationally stable, endowing it with increased potency and longevity. The conjugation of JA and l-Ile in plants is catalysed by the adenosine triphosphate (ATP)-dependent ligase Jar1 (Fig. **1b**), a member of the ANL superfamily<sup>5</sup>. ANL enzymes generate acyl-adenylate (acyl-AMP) intermediates that undergo substitution with various nucleophiles. For example, acyl-CoA synthetases (ACSs) are common ANL enzymes generating thioesters, which can be coupled to an amine by a secondary *N*-acyltransferase. Jar1 is an amide bond synthetase (ABS), a rarer subclass of the ANL superfamily which accept amine nucleophiles directly without requiring an additional partner enzyme (Supplementary Fig. **1**)<sup>5,6</sup>.

**Fig. 1: Biosynthesis of coronatine (COR) and jasmonyl-l-isoleucine (JA-Ile).**

---

 **figure1**

**a**, Bacterial CfaL enzymes are predicted to ligate coronafacic acid (CFA) **1** with coronamic acid (CMA) **2** or l-amino acids to generate phytotoxins, including COR **3** and CFA-Ile **4**. **b**, In plants, the enzyme Jar1 ligates jasmonic acid (JA) and l-isoleucine to produce JA-Ile epimers (*3R,7S*)-**6** and (*3R,7S*)-**6**. **c**, Polyketide synthase (PKS) assembly of **1** from succinic semialdehyde. PKS consists of acyl-carrier protein (ACP), acyl-transferase (AT), keto-synthase (KS), dehydratase (DH), enoyl-reductase (ER), keto-reductase (KR) and thioesterase (TE) domains. **d**, Nonribosomal peptide synthetase (NRPS)-mediated biosynthesis of **2** in *P. syringae*. NRPS consists of adenylation (A), thiolation (T) and thioesterase (TE) domains.

[Full size image](#)

CFA is assembled by a type I polyketide synthase via the cyclization of a  $\beta$ -ketothioester intermediate (Fig. 1c)<sup>1,7,8,9</sup>. The biosynthesis of CMA occurs via a nonribosomal peptide synthetase (NRPS)-mediated cryptic chlorination and cyclization (Fig. 1d)<sup>10,11,12</sup>. A putative ligase (CfaL) within the *Pseudomonas syringae* COR biosynthetic gene cluster is predicted to couple CFA and CMA to form COR (Fig. 1a, Supplementary Fig. 2a and Supplementary Table 1). Previous attempts to characterize this ligase have been unsuccessful<sup>13</sup>. Other plant pathogens possess a COR-like biosynthetic gene cluster<sup>14,15</sup>, including *Streptomyces scabies* which has a putative CfaL and CFA biosynthetic genes, but no CMA pathway (Supplementary Fig. 2b and Supplementary Table 1)<sup>15</sup>. Consequently, this strain produces predominantly CFA-l-Ile 4, along with smaller quantities of l-Val and l-*allo*-Ile adducts, which have also been detected from *P. syringae*<sup>16,17,18</sup>. As CFA must be coupled to an amino acid to elicit biological activity, we sought to characterize the key CfaL enzymes to enable new routes to COR-like phytotoxins as potential herbicides<sup>19</sup>. We were also interested in exploring the relationship between the bacterial CfaL and the functionally related plant ligase Jar1.

As well as being fundamental in nature, amide formation is one of the most widely used synthetic transformations. Although coupling acids and amines is relatively simple, it often requires three steps, protect–couple–deprotect, to install each amide. Stoichiometric quantities of expensive and deleterious coupling reagents are typically required and purification can be problematic<sup>20</sup>. While some progress has been made in the development of chemocatalytic methods for amide synthesis, these have not been widely adopted<sup>21,22,23,24,25,26,27</sup>. Consequently, there is interest in the development of enzymatic alternatives<sup>6,20,28,29,30,31</sup>. In this work, we characterize CfaL ligases and demonstrate how they are highly versatile biocatalysts for the synthesis of ubiquitous amides. Additionally, using structure-guided mutagenesis, we generate improved ligases providing more sustainable, alternative routes for production of pharmaceuticals, agrochemicals and other valuable materials.

## Characterization of the CfaL family

Overproduction of the CfaL from *P. syringae* (PsCfaL) in *Escherichia coli* resulted in only trace amounts of active enzyme. However, assays demonstrated that PsCfaL catalyses the ATP-dependent coupling of l-isoleucine and CFA, obtained from acid hydrolysis of coronatine, confirming the function of CfaL for the first time (Supplementary Figs. 3 and 4). The low quantity of PsCfaL available prevented full characterization, and so alternative CfaL homologues were explored. In addition to the putative *S. scabies* ligase (SsCfaL), other candidates located within putative COR-like

clusters were selected from BLAST analysis (Supplementary Fig. 2). Of these, PbCfaL from *Pectobacterium brasiliense* and AlCfaL from *Azospirillum lipoferum* were chosen for characterization, as they are predicted to be more amenable to crystallization (Supplementary Table 2). While *P. brasiliense* is a well-known plant pathogen<sup>32,33</sup>, *A. lipoferum* is a root-dwelling, nitrogen-fixing plant symbiont that is not known to produce coronatine<sup>34</sup>.

SsCfaL was overproduced in *E. coli* (Supplementary Fig. 5) and assays with synthetic ( $\pm$ )-CFA<sup>19</sup>, l-isoleucine and ATP showed the direct formation of CFA-l-Ile 4 via a CFA-AMP intermediate (Supplementary Fig. 6), confirming CfaL is an ABS enzyme. In addition, SsCfaL also accepted the aromatic CFA variant 7 which when coupled to l-Ile forms coronalone, a simplified synthetic COR analogue with promising herbicidal activity<sup>35</sup>. Given adenylation occurs in the absence of amine substrate, the rate of adenylation can be measured in isolation (Extended Data Table 1 and Supplementary Fig. 7). The rate of adenylation ( $k_{cat}$ ) was greatest with ( $\pm$ )-CFA but the aromatic analogue 7 was found to have a lower Michaelis constant,  $K_m$ . Both (3*R*,7*R*)- and (3*S*,7*S*)-enantiomers of *trans*-JA 5 were also accepted, albeit at a lower level than CFA or 7, with a preference towards the natural (3*R*,7*R*)-5 stereoisomer. This indicates that, despite low amino acid sequence similarity (16%), the bacterial SsCfaL and plant Jar1 both catalyse the ligation of JA with l-Ile (Supplementary Fig. 8). Reactions with deactivated enzyme confirmed that the adenylation of 5 and the subsequent reaction with l-Ile are both enzyme-catalysed. Samples of *trans*-JA (Extended Data Table 1) contain a minor amount of the less stable *cis* epimer, owing to facile C7-epimerization (Fig. 1b)<sup>4</sup>. To explore the stereoselectivity of SsCfaL further, all four stereoisomers of configurationally stable 7-methyl-jasmonic acid were synthesized (Supplementary Fig. 9). Initially the *trans* and *cis* diastereoisomers of 7-methyl-jasmonic acid were separated and tested as a racemic mixture. However, these were poor substrates for SsCfaL, hence the resolution of all four stereoisomers was not carried out. The selectivity of SsCfaL was further tested by incubating CFA 1 or ( $\pm$ )-jasmonic acid 5 with 21 proteinogenic amino acids, resulting in a wide range of amino acid conjugates (Supplementary Figs. 10 and 11). Hydrophobic amino acids such as l-isoleucine and l-valine were preferred by SsCfaL which reflects the COR-like metabolites isolated from *S. scabies*<sup>16</sup>. No activity was seen with d-amino acids, or with primary amines and dipeptides (Supplementary Fig. 12). AlCfaL and PbCfaL, expressed from codon optimized synthetic genes, were seen to be functionally similar to SsCfaL, although with lower activities (Supplementary Fig. 13).

## The structure of a CfaL ligase

Crystallography trials revealed that only PbCfaL yielded crystals of sufficient quality for structural studies. A PbCfaL structure in the adenylation conformation was solved

to 2 Å resolution (Fig. 2), which is consistent with the ANL superfamily. Despite sharing low sequence identity (<20%), PbCfaL showed substantial structural similarity (>70%) to several other ANL ligases from a variety of organisms, including bacterial benzoate CoA ligases and firefly luciferases (Fig. 2b, Supplementary Table 3). By contrast, PbCfaL shares very little structural similarity with the catalytically equivalent Jar1 (30%), suggesting that the two enzymes evolved independently (Supplementary Fig. 14). Plants are known to possess other ACSs that do share high structural homology with PbCfaL (Supplementary Table 3). However, Jar1 and related plant acyl-AMP forming enzymes that conjugate salicylate or indole-6-acetic acid (IAA) with amino acids appear to have evolved separately and specifically for plant hormone signalling<sup>36</sup>.

**Fig. 2: Structure of PbCfaL.**

---

 **figure2**

**a**, Main image, X-ray crystal structure of PbCfaL (2 Å) in the ‘open’ or ‘adenylation’ conformation (PDB ID, 7A9I). PbCfaL has a large N-terminal region (residues 1–403), shown in blue, and a flexible C-terminal region (residues 403–516), shown in red. The boxed region and magnified inset highlight the active site region, which is

shown with **7** co-crystallized; labelled residues are conserved between all four CfaLs in this study. **7** lies 3.6 Å from a conserved tryptophan residue (W220), which probably helps to align the substrate via π–π stacking interactions. **b**, PbCfaL superimposed onto McbA (gold; PDB ID, 6SQ8)<sup>29</sup>, the closest structural homologue to PbCfaL. In the ‘adenylation’ state, the two structures show high levels of similarity. In this state the carboxylic acid binding site (shown) is located between the N-terminal and the flexible C-terminal regions and is solvent accessible. **c**, PbCfaL superimposed on McbA in the ‘closed’ state (also referred to as the ‘thiolation’ state in related ACS enzymes). Like all ANLs, the C-terminal region of McbA undergoes a large rotation (direction indicated by dashed red arrow) to lie on top of the carboxylic acid binding site, trapping the adenylated intermediate before amine attack. The more rigid N-terminal region does not substantially change conformation. We would expect the C-terminal region of PbCfaL to undergo a similar rotation during catalysis. Structural alignment was performed with Chimera (version 1.14) MatchMaker.

[Full size image](#)

The structure of PbCfaL is composed of a large N-terminal domain (residues 1–403) and a smaller, flexible C-terminal domain (residues 403–516, Fig. 2). As with other members of the ANL superfamily, it is likely that the C-terminal domain undergoes a large rotation following acyl-adenylate formation to close off the acyl binding pocket and form the amino acid binding site (closed conformation) (Fig. 2). Co-crystallography of PbCfaL with **7** revealed that the extremely solvent-accessible acyl binding pocket lies between the two domains (Fig. 2, Supplementary Fig. 15). Sequence alignment between the CfaLs in this study showed a small number of conserved residues (Supplementary Fig. 16), with only W220 likely to make direct contact with **7**, probably aligning the carboxylic acid via π–π stacking, explaining the higher binding affinity of **7** versus CFA (Fig. 2). Other conserved residues around **7** probably define the width and depth of the binding pocket. When aligned with the most structurally similar proteins from the Protein Data Bank (PDB; Supplementary Fig. 17) there are few conserved sequences, the most similarity occurring in the ATP binding SSGTTG motif (residues 168–173)<sup>37</sup>. Despite many attempts, determination of a PbCfaL structure in the closed conformation with AMP and amino acid bound could not be achieved.

## CfaL substrate scope and engineering

We next explored if the synthetic scope of CfaL enzymes could be extended towards other amide targets. The CfaL enzymes were found to possess extremely broad substrate tolerance, accepting a variety of aryl and heteroaryl carboxylic acids **8–29** as well as aliphatic carboxylic acids **30–46**, including several chiral compounds **39–46** (Fig. 3, Extended Data Table 2). Several acyl-donor substrates possessed other reactive

functionalities, such as electrophilic ketones (**11**, **35**, **43**, **45**), alkenes (**33**), as well as nucleophilic alcohol (**40**, **44**) or amine groups (**13**, **17**, **18**, **19**, **46**), which would require protecting for traditional coupling chemistries, but do not interfere with the enzymatic ligation to amino acid acceptor substrates. In addition to proteinogenic amino acids (Fig. [4a](#), Extended Data Table [3](#)), CfaL enzymes also accept a wide range of non-proteinogenic amino acids **47–61**, including common pharmaceutical building blocks, with a preference for hydrophobic amino acids (Fig. [4b,c](#), Extended Data Table [4](#)). Although polar, particularly charged, amino acids are not well accepted by CfaL, both L-2,4-diaminobutyrate **47** and L-ornithine **48** can be selectively acylated at the  $\alpha$ -amino group, obviating the need for protection of the side-chain amino group (Fig. [4c](#), Supplementary Fig. [17](#)).

**Fig. 3: Carboxylic acid substrate scope.**



**a**, Diverse structures of carboxylic acid (donor) substrates assayed with l-Ile and CfaL enzymes. **b**, Percentage conversion for ligation of carboxylic acids **8–46** with l-Ile catalysed by CfaL enzymes (column headings). Assays were carried out with wild-type and engineered CfaL enzymes (25  $\mu$ M), carboxylic acids **8–46** (2 mM) and l-Ile (5 mM). Conversion to amide products was determined by HPLC analysis following 20 h incubation. Actual conversion values and errors can be found in Extended Data Table [2](#).

[Full size image](#)

**Fig. 4: Amino acid substrate scope.**

 [figure4](#)

**a**, Percentage conversion for ligation of carboxylic acid **9** with acceptor proteinogenic amino acids (rows). **b**, Percentage conversion for ligation of **9** with acceptor non-proteinogenic amino acids. a-Ile = *allo*-isoleucine. **c**, Structures of non-proteinogenic amino acids. **d**, Reversed-phase (RP)-HPLC trace of ligation product of l-Dab **47** (green) and **9** (*m*-methylbenzoate, red) catalysed by SsCfaL, compared to HPLC traces of synthesized standards of the two possible products **62** and **63**. Product of the enzymatic reaction (bottom trace) shows selective acylation of the  $\alpha$ -amino group to give amide **62**. All assays were carried out with wild-type and engineered CfaL enzymes (5  $\mu$ M), carboxylic acid **9** (1 mM) and amino acids (2 mM). Conversion to amide products was determined by RP-HPLC analysis following 20 h incubation. Actual conversion values and errors can be found in Extended Data Tables [3](#) and [4](#).

[Full size image](#)

In general, SsCfaL and AlCfaL both performed better than PbCfaL (Figs. 3 and 4, Extended Data Tables 2–4), which was found to be less thermally stable and frequently precipitated during the reaction timescale (Extended Data Fig. 1). On the basis of our crystallographic studies, we sought to improve the activity and stability of PbCfaL via rational, structure-guided mutagenesis. Sequence comparison between the four CfaL enzymes (Supplementary Fig. 16) identified few obvious distinctions. However, one noticeable difference was found on the flexible hinge-region linking the N- and C-terminal domains (at position 395, Extended Data Fig. 2a). This position is solvent-exposed and likely to be involved in the conformational changes required to shift between the adenylation (open) and amidation (closed) states of CfaL. The large, charged arginine residue that is located in this position of PbCfaL is orientated out from the enzyme, while the same position in the other CfaLs and ANLs included in the sequence alignment (Supplementary Fig. 16) is occupied by a small and uncharged glycine. A PbCfaL(R395G) mutant showed increased activity against the panel of carboxylic acids and some amino acid substrates (Figs. 3 and 4, Extended Data Tables 2–4). An X-ray crystal structure of this mutant was determined, which revealed no overall structural changes (Extended Data Fig. 2a). However, the melting temperature ( $T_m$ ) of PbCfaL(R395G) increased by 5 °C relative to the wild type, suggesting that the replacement of this solvent-accessible, charged R395 is beneficial for stability (Extended Data Fig. 1).

A subsequent double mutant, PbCfaL(R395G/A294P), showed a further increased  $T_m$  and slightly improved activity (Figs. 3 and 4, Extended Data Table 2–4). The location of this second mutation is within a highly conserved ATP binding loop (G289–L297) that is significantly larger in PbCfaL than in other related structures, and which may partially occlude the ATP binding site (Extended Data Fig. 2b). The proline found at this location in SsCfaL, PbCfaL and several other structurally similar ligases may aid in rotating this loop out of the binding site (Supplementary Fig. 16). Using these two PbCfaL mutants, which no longer precipitate during the reaction, we were able to substantially improve the conversions of both the panel of carboxylic acid and amino acid substrates (Figs. 3 and 4), demonstrating that minimal structure-guided mutagenesis can be used to engineer improved CfaL variants.

## Synthetic applications of CfaL

To demonstrate the synthetic utility of CfaL, we sought to establish preparative-scale ligation reactions. Accordingly, conditions were optimized for the ligation of carboxylic acid **10** and 1-Ile (Fig. 5a). Reaction of **10** (at 15 mM concentration) with PbCfaL(R395G/A294P) cell free lysate afforded amide **64** in near quantitative conversion as determined by high-performance liquid chromatography (HPLC). The

reaction mixture was subjected to a simple solvent extraction, providing 1.48 g of crude **64** from 400 ml of reaction mixture, which would be sufficiently pure (>92% purity by NMR, Extended Data Fig. [3](#)) for further synthetic derivatization. Purification of the extract by column chromatography provided 1.37 g of pure **64** in 87% isolated yield (Fig. [5a](#)). To avoid the use of stoichiometric quantities of the expensive co-factor ATP, we repeated the ligation of **10** and l-Ile at the same scale, omitting ATP and instead introducing an ATP recycling system consisting of a polyphosphate (PolyP) kinase enzyme (CHU)<sup>[38](#)</sup> and an inexpensive PolyP phosphate donor (Extended Data Fig. [3](#)). Although the isolated yield was lower in this case (52%), there is further scope for optimization. While CfaL cell lysate shows good activity for up to 12 h, we sought to improve enzyme stability/longevity through immobilization of CfaL in the form of a cross-linked enzyme aggregate (CLEA)<sup>[39](#)</sup>. PbCfaL(R395G/A294P) CLEAs were shown to retain activity over an extended period of five days, and could be isolated and recycled in five sequential ligation reactions (Extended Data Fig. [4a](#)). Purified PbCfaL(R395G/A294P) was also shown to tolerate several solvents, including MeOH, ethylene glycol and the widely used ‘green’ solvent 2-methylTHF (Extended Data Fig. [4b](#)).

**Fig. 5: Synthetic potential of CfaL enzymes.**

---

 figure5

**a**, Amides including pharmaceutical scaffolds synthesized by CfaL enzymes. **b**, Comparison of percentage conversion for CfaL enzymes in the synthesis of **65–70**. **c**, Kinetic resolution of racemic carboxylic acids (donor). Absolute configuration and diastereoisomeric ratios (d.r.) were determined by RP-HPLC using synthetic standards. Inset, the HPLC chromatogram for (*S*)-**71** formed in the kinetic resolution of racemic ibuprofen (**41**) with l-Ile and AlCfaL ( $E = 94$ ). **d**, Comparison of the enantioselectivities of the different enzymes. Values of  $E = 15–30$  are considered moderate–good,  $E > 30$  are excellent<sup>43</sup>. For conversions  $<30\%$  the calculation of  $E$  is unreliable, so the values were not determined (ND). **e**, Kinetic resolution of racemic amino acid (acceptor) **57** (2 mM) with acid **9** (1 mM) and AlCfaL (5  $\mu$ M) ( $E > 200$ ) following 20 h incubation. The yield reported is based on **9** which equates to a yield of 33% based on **57** (2 equiv. used). Inset, chiral HPLC analysis of the amide product showing a single enantiomer, (*S*)-**77**. Enantiomeric ratio (e.r.) values determined by chiral HPLC. <sup>a</sup>Isolated yield preparative-scale synthesis of **64** with PbCfaL(R395G/A294P) lysate **10** (15 mM), l-Ile (45 mM) and ATP (36 mM) incubated for 24 h. <sup>b</sup>Isolated yields of about 100-mg-scale reactions catalysed by

SsCfaL cell lysate with carboxylic acid (5 mM), amine (15 mM) and ATP (15 mM) following 24 h incubation. <sup>c</sup>Conversions determined from HPLC peak area ratios, following assays including CfaL enzymes (25 µM), carboxylic acids (2 mM) and amino acid (6 mM) incubated for 20 h. <sup>d</sup>*E* value was calculated from average d.r. or e.r. values, as described previously<sup>43</sup>. Percentage conversions and d.r. values represent means where *n* = 3, error denotes s.d.

[Full size image](#)

To further demonstrate the synthetic potential of CfaL, a series of pharmaceutical-relevant scaffolds were prepared in excellent yields (Fig. 5a,b and Supplementary Fig. 18). For example, amides 65 and 67 were prepared in >70% isolated yields at around 100 mg scale. Furthermore, ligations of cinnamic acid 33 and indole carboxylic acids 26 (Fig. 3) with cyclopropyl amino acids 56 and l-Leu, respectively (Fig. 4), produced amides 68 and 69, which are precursors for the manufacture of promising SARS-CoV-2 protease inhibitors, including PF-07304814 (Pfizer; in phase I clinical trials) (Fig. 5a,b and Supplementary Fig. 18)<sup>40,41</sup>. Similarly, ligation of the thiazole carboxylic acid 29 and *O*-methyl-l-serine 49 provided amide 70; this is a key component of oprozomib, which is in phase II clinical trials for treatment of multiple myeloma<sup>42</sup>. Probing the limits of potential CfaL-reaction scope, we found that the mutant PbCfaL(R395G/A294P) also allowed the generation of amide precursors required for the synthesis of the antiviral telaprevir and the anti-cancer agent bortezomib (Extended Data Fig. 5). These products were only produced in low quantities, but with further engineering it may be possible to synthesize these precursors at higher levels. Overall, these reactions (Fig. 5) clearly demonstrate how structurally diverse carboxylic acids can be combined with proteinogenic and synthetic amino acids to produce pharmaceutically important compounds.

The potential of CfaL for use in kinetic resolution of racemic synthetic carboxylic acids was also investigated (Fig. 5c,d). Notably, racemic ibuprofen could be resolved, with excellent enantioselectivity (*E* = 94)<sup>43</sup> leading to the biologically active (*S*)-ibuprofen-l-Ile amide 71. Amide conjugates of ibuprofen and related NSAIDs with amino acids have been explored extensively for applications as prodrugs and/or hydrogel-based nanomedicine<sup>44,45</sup>. Five other racemic acids were subjected to kinetic resolution, affording amides 72–76, with modest *E* values (Fig. 5c,d and Supplementary Fig. 19). In the case of amide 73, the mutant PbCfaL(R395G) was superior to any of the wild-type CfaLs, illustrating how protein engineering could be used to achieve more effective kinetic resolutions. Finally, we sought to exploit the high selectivity of CfaL for l-amino acids to effect the kinetic resolution of racemic amino acids. Amino acid 57 was selected as this is a common pharmaceutical building block, which would normally require multi-step asymmetric synthesis or laborious resolution and protection before acylation or peptide coupling. As anticipated, the

reaction between carboxylic acid **9** and racemic amino acid **57** proceeds with excellent enantioselectivity ( $E > 200$ ), with none of the *R*-configured enantiomer evident in chiral HPLC when SsCfaL or AlCfaL was used (Fig. [5e](#), Supplementary Fig. [20](#)). This demonstrates how racemic carboxylic acid and racemic amino acids can be resolved during amide bond synthesis, using CfaL, avoiding more laborious asymmetric synthesis or traditional resolution procedures, and the need for protective group manipulations.

## Discussion

The results presented here demonstrate the role of CfaL enzymes in biosynthesis of the important coronatine family of phytotoxins. BLAST analysis reveals that CfaL-like ligases appear in a large number of distinct COR-like clusters from across a broad range of microorganisms, including bacteria where COR-like phytotoxins have not been observed, suggesting that CfaLs and the biosynthesis of COR-like phytotoxins are widespread. CfaLs can also catalyse ligation of JA with Ile to generate the plant hormone JA-Ile in an identical fashion to the plant ligase Jar1. The lack of sequence and structural similarity between the CfaL and Jar1 suggests that the two enzymes have evolved largely independently in bacteria and in plants to perform very similar reactions. In addition to potential agrochemical applications, the CfaL family of enzymes can be used to produce a wide range of pharmaceutically relevant amides. Used in combination with improving ATP recycling techniques<sup>[46,47](#)</sup>, these enzymes could become powerful synthetic tools offering major advantages over other biocatalysts developed for amide synthesis. For example, the combination of ACS and *N*-acyltransferase enzymes have been investigated for amide synthesis<sup>[48](#)</sup>. However, large numbers of ACS and *N*-acyltransferase enzymes had to be screened to find pairs of enzymes with matching selectivity<sup>[48](#)</sup>. In addition to low substrate scope, this system also requires use of two expensive co-factors as well as engineering of two enzymes for further optimization, rather than just one. Other reports describe the use of standalone NRPS adenylation domains to synthesize amides<sup>[30,49](#)</sup>. In these examples only the carboxylic acid activation step is directly enzyme catalysed, the subsequent amidation proceeds spontaneously, requiring a large excess (about 100 equiv.) of the amine, which is not viable for many syntheses. CfaLs directly catalyse both steps and can therefore utilize acids and amines with more efficient stoichiometry. Taken together our results show that CfaLs have potential for the synthesis of diverse range of important amide products, offering clear advantages over traditional synthetic methods and other biocatalytic approaches.

## Reporting summary

Further information on research design is available in the [Nature Research Reporting Summary](#) linked to this paper.

## Data availability

Nucleotide sequences for the mutants generated as part of this study are available in [Supplementary Information](#). Other nucleotide sequences for the enzymes used in this study were obtained from GenBank, and their accession numbers are provided within the paper or in [Supplementary Information](#). The original materials and data that support the findings of this study are either available within the paper or are available from the corresponding author upon reasonable request. Crystallographic coordinates of wild type and mutant PbCfaL have been deposited in the Protein Data Bank as [7A9I](#) (wild type) and [7A9J](#) (R395G mutant).

## References

1. 1.

Rangaswamy, V., Jiralerspong, S., Parry, R. & Bender, C. L. Biosynthesis of the *Pseudomonas* polyketide coronafacic acid requires monofunctional and multifunctional polyketide synthase proteins. *Proc. Natl Acad. Sci. USA* **95**, 15469–15474 (1998).

[ADS](#) [CAS](#) [Google Scholar](#)

2. 2.

Ullrich, M. & Bender, C. L. The biosynthetic gene cluster for coronamic acid, an ethylcyclopropyl amino acid, contains genes homologous to amino acid-activating enzymes and thioesterases. *J. Bacteriol.* **176**, 7574–7586 (1994).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

3. 3.

Staswick, P. E. & Tiryaki, I. The oxylipin signal jasmonic acid is activated by an enzyme that conjugates it to isoleucine in *Arabidopsis*. *Plant Cell* **16**, 2117–2127 (2004).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

4. 4.

Fonseca, S. et al. (+)-7-iso-Jasmonoyl-l-isoleucine is the endogenous bioactive jasmonate. *Nat. Chem. Biol.* **5**, 344–350 (2009).

[CAS](#) [Google Scholar](#)

5. 5.

Westfall, C. S. et al. Structural basis for prereceptor modulation of plant hormones by GH3 proteins. *Science* **336**, 1708–1711 (2012).

[ADS](#) [CAS](#) [Google Scholar](#)

6. 6.

Winn, M., Richardson, S. M., Campopiano, D. J. & Micklefield, J. Harnessing and engineering amide bond forming ligases for the synthesis of amides. *Curr. Opin. Chem. Biol.* **55**, 77–85 (2020).

[CAS](#) [Google Scholar](#)

7. 7.

Parry, R. J., Jiralerspong, S., Mhaskar, S., Alemany, L. & Willcott, R. Investigations of coronatine biosynthesis. Elucidation of the mode of incorporation of pyruvate into coronafacic acid. *J. Am. Chem. Soc.* **118**, 703–704 (1996).

[CAS](#) [Google Scholar](#)

8. 8.

Tao, T. & Parry, R. J. Determination by enantioselective synthesis of the absolute configuration of CPE, a potential intermediate in coronatine biosynthesis. *Org. Lett.* **3**, 3045–3047 (2001).

[CAS](#) [Google Scholar](#)

9. 9.

Strieter, E. R., Koglin, A., Aron, Z. D. & Walsh, C. T. Cascade reactions during coronafacic acid biosynthesis: elongation, cyclization, and functionalization during Cfa7-catalyzed condensation. *J. Am. Chem. Soc.* **131**, 2113–2115 (2009).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

10. 10.

Parry, R. J., Lin, M. T., Walker, A. E. & Mhaskar, S. Biosynthesis of coronatine: investigations of the biosynthesis of coronamic acid. *J. Am. Chem. Soc.* **113**, 1849–1850 (1991).

[CAS](#) [Google Scholar](#)

11. 11.

Vaillancourt, F. H., Yeh, E., Vosburg, D. A., O'Connor, S. E. & Walsh, C. T. Cryptic chlorination by a non-haem iron enzyme during cyclopropyl amino acid biosynthesis. *Nature* **436**, 1191–1194 (2005).

[ADS](#) [CAS](#) [Google Scholar](#)

12. 12.

Kelly, W. L. et al. Characterization of the aminocarboxycyclopropane-forming enzyme CmaC. *Biochemistry* **46**, 359–368 (2007).

[CAS](#) [Google Scholar](#)

13. 13.

Rangaswamy, V. et al. Expression and analysis of coronafacate ligase, a thermoregulated gene required for production of the phytotoxin coronatine in *Pseudomonas syringae*. *FEMS Microbiol. Lett.* **154**, 65–72 (1997).

[CAS](#) [Google Scholar](#)

14. 14.

Slawiak, M. & Lojkowska, E. Genes responsible for coronatine synthesis in *Pseudomonas syringae* present in the genome of soft rot bacteria. *Eur. J. Plant Pathol.* **124**, 353–361 (2009).

[CAS](#) [Google Scholar](#)

15. 15.

Bignell, D. R. D. et al. *Streptomyces scabies* 87–22 contains a coronafacic acid-like biosynthetic cluster that contributes to plant-microbe interactions. *Mol. Plant Microbe Interact.* **23**, 161–175 (2010).

[CAS](#) [Google Scholar](#)

16. 16.

Fyans, J. K., Altowairish, M. S., Li, Y. & Bignell, D. R. Characterization of the coronatine-like phytotoxins produced by the common scab pathogen *Streptomyces scabies*. *Mol. Plant Microbe Interact.* **28**, 443–454 (2015).

[CAS](#) [Google Scholar](#)

17. 17.

Mitchell, R. E. & Frey, E. J. Production of *N*-coronafacyl-L-amino-acid-analogs of coronatine by *Pseudomonas syringae* pv *Atropurpurea* in liquid cultures supplemented with L-amino acids. *J. Gen. Microbiol.* **132**, 1503–1507 (1986).

[CAS](#) [Google Scholar](#)

18. 18.

Mitchell, R. E. & Ford, K. L. Chlorosis-inducing products from *Pseudomonas syringae* pathovars: new *N*-coronafacyl compounds. *Phytochemistry* **49**, 1579–1583 (1998).

[CAS](#) [Google Scholar](#)

19. 19.

Littleson, M. M. et al. Scalable total synthesis and comprehensive structure-activity relationship studies of the phytotoxin coronatine. *Nat. Commun.* **9**, 1105 (2018).

[ADS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

20. 20.

Sabatini, M. T., Boulton, L. T., Sneddon, H. F. & Sheppard, T. D. A green chemistry perspective on catalytic amide bond formation. *Nat. Catal.* **2**, 10–17 (2019).

[CAS](#) [Google Scholar](#)

21. 21.

Sabatini, M. T., Boulton, L. T. & Sheppard, T. D. Borate esters: simple catalysts for the sustainable synthesis of complex amides. *Sci. Adv.* **3**, e1701028 (2017).

[ADS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

22. 22.

Krause, T., Baader, S., Erb, B. & Gooßen, L. J. Atom-economic catalytic amide synthesis from amines and carboxylic acids activated *in situ* with acetylenes. *Nat. Commun.* **7**, 11732 (2016).

[ADS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

23. 23.

Stephenson, N. A., Zhu, J., Gellman, S. H. & Stahl, S. S. Catalytic transamidation reactions compatible with tertiary amide metathesis under ambient conditions. *J. Am. Chem. Soc.* **131**, 10003–10008 (2009).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

24. 24.

Allen, C. L., Atkinson, B. N. & Williams, J. M. J. Transamidation of primary amide with amines using hydroxylamine hydrochloride as an inorganic catalyst. *Angew. Chem. Int. Edn* **51**, 1383–1386 (2012).

[CAS](#) [Google Scholar](#)

25. 25.

Al-Zoubi, R. M., Marion, O. & Hall, D. G. Direct and waste-free amidations and cycloadditions by organocatalytic activation of carboxylic acids at room temperature. *Angew. Chem. Int. Edn* **47**, 2876–2879 (2008).

[CAS](#) [Google Scholar](#)

26. 26.

Noda, H., Furutachi, M., Asada, Y., Shibasaki, M. & Kumagai, N. Unique physicochemical and catalytic properties dictated by the B<sub>3</sub>NO<sub>2</sub> ring system. *Nat. Chem.* **9**, 571–577 (2017).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

27. 27.

Scheidt, K. Amide bonds made in reverse. *Nature* **465**, 1020–1022 (2010).

[ADS](#) [CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

28. 28.

Goswami, A. & Van Lanen, S. G. Enzymatic strategies and biocatalysts for amide bond formation: tricks of the trade outside of the ribosome. *Mol. Biosyst.* **11**, 338–353 (2015).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

29. 29.

Petchey, M. et al. The broad aryl acid specificity of the amide bond synthetase McbA suggests potential for the biocatalytic synthesis of amides. *Angew. Chem. Int. Edn* **57**, 11584–11588 (2018).

[CAS](#) [Google Scholar](#)

30. 30.

Wood, A. J. L. et al. Adenylation activity of carboxylic acid reductases enables the synthesis of amides. *Angew. Chem. Int. Edn* **56**, 14498–14501 (2017).

[CAS](#) [Google Scholar](#)

31. 31.

Pattabiraman, V. R. & Bode, J. W. Rethinking amide bond synthesis. *Nature* **480**, 471–479 (2011).

[ADS](#) [CAS](#) [Google Scholar](#)

32. 32.

Pérombelon, M. C. M. Potato diseases caused by soft rot *Erwinias*: an overview of pathogenesis. *Plant Pathol.* **51**, 1–12 (2002).

[Google Scholar](#)

33. 33.

Bell, K. S. et al. Genome sequence of the enterobacterial phytopathogen *Erwinia carotovora* subsp. *atroseptica* and characterization of virulence factors. *Proc. Natl Acad. Sci. USA* **101**, 11105–11110 (2004).

[ADS](#) [CAS](#) [Google Scholar](#)

34. 34.

Bottini, R., Fulchieri, M., Pearce, D. & Pharis, R. P. Identification of gibberellins A<sub>1</sub>, A<sub>3</sub> and iso-A<sub>3</sub> in cultures of *Azospirillum lipoferum*. *Plant Physiol.* **90**, 45–47 (1989).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

35. 35.

Schüler, G. et al. Coronalon: a powerful tool in plant stress physiology. *FEBS Lett.* **563**, 17–22 (2004).

[Google Scholar](#)

36. 36.

Shockley, J. M., Fulda, M. S. & Browse, J. *Arabidopsis* contains a large superfamily of acyl-activating enzymes. Phylogenetic and biochemical analysis reveals a new class of acyl-coenzyme A synthetases. *Plant Physiol.* **132**, 1065–1076 (2003).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

37. 37.

Stuhlsatz-Krouper, S. M., Bennett, N. E. & Schaffer, J. E. Substitution of alanine for serine 250 in the murine fatty acid transport protein inhibits long chain fatty acid transport. *J. Biol. Chem.* **273**, 28642–28650 (1998).

[CAS](#) [Google Scholar](#)

38. 38.

Nocek, B. P. et al. Structural insights into substrate selectivity and activity of bacterial polyphosphate kinases. *ACS Catal.* **8**, 10746–10760 (2018).

[CAS](#) [Google Scholar](#)

39. 39.

Sheldon, R. A. Cross-linked enzyme aggregates (CLEAs): stable and recyclable biocatalysts. *Biochem. Soc. Trans.* **35**, 1583–1587 (2007).

[CAS](#) [Google Scholar](#)

40. 40.

Zhang, L. et al.  $\alpha$ -Ketoamides as broad-spectrum inhibitors of coronavirus and enterovirus replication: structure-based design, synthesis, and activity assessment. *J. Med. Chem.* **63**, 4562–4578 (2020).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

41. 41.

Boras, B. et al. Discovery of a novel inhibitor of coronavirus 3CL protease for the potential treatment of COVID-19. Preprint at <https://www.biorxiv.org/content/10.1101/2020.09.12.293498v3> (2020).

42. 42.

Zhou, H.-J. et al. Design and synthesis of an orally bioavailable and selective peptide epoxyketone proteasome inhibitor (PR-047). *J. Med. Chem.* **52**, 3028–3038 (2009).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

43. 43.

Chen, C.-S., Fujimoto, Y., Girdaukas, G. & Sih, C. J. Quantitative analyses of biochemical kinetic resolutions of enantiomers. *J. Am. Chem. Soc.* **104**, 7294–7299 (1982).

[CAS](#) [Google Scholar](#)

44. 44.

Jervis, P. J., Amorim, C., Pereira, T., Martins, J. A. & Ferreira, P. M. T. Exploring the properties and potential biomedical applications of NSAID-capped peptide hydrogels. *Soft Matter* **16**, 10001–10012 (2020).

[ADS](#) [CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

45. 45.

Tiwari, A. D. et al. Microwave assisted synthesis and QSAR study of novel NSAID acetaminophen conjugates with amino acid linkers. *Org. Biomol. Chem.* **12**, 7238–7249 (2014).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

46. 46.

Andexer, J. N. & Richter, M. Emerging enzymes for ATP regeneration in biocatalytic processes. *ChemBioChem* **16**, 380–386 (2015).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

47. 47.

Strohmeier, G. A., Eiteljorg, I. C., Schwarz, A. & Winkler, M. Enzymatic one-step reduction of carboxylates to aldehydes with cell-free regeneration of ATP and NADPH. *Chem. Eur. J.* **25**, 6119–6123 (2019).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

48. 48.

Philpott, H. K., Thomas, P. J., Tew, D., Fuerst, D. E. & Lovelock, S. L. A versatile biosynthetic approach to amide bond formation. *Green Chem.* **20**, 3426–3431 (2018).

[CAS](#) [Google Scholar](#)

49. 49.

Hara, R., Hirai, K., Suzuki, S. & Kino, K. A chemoenzymatic process for amide bond formation by an adenylating enzyme-mediated mechanism. *Sci. Rep.* **8**, 2950 (2018).

[ADS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

[Download references](#)

## Acknowledgements

We thank the BBSRC (grants BB/K002341/1 and BB/N023536/1) and Syngenta for funding. F.W. was supported by the China Scholarship Council (grant no. 201806155100) and L.B. was funded by the Deutsche Forschungsgemeinschaft (DFG, grant BE 7054/1). The Michael Barber Centre for Collaborative Mass Spectrometry provided access to MS instrumentation. We also thank J. Vincent and N. Mulholland (Syngenta) for helpful discussions in the early stages of the project, and N. J. Turner (University of Manchester) for kindly providing the CHU plasmid. We also thank Diamond Light Source for beamtime access on i03 and i04-1 (proposal mx17773-56 and 76).

## Author information

### Author notes

#### 1. Fanghua Wang

Present address: School of Food Science and Engineering, South China University of Technology, Guangzhou, China

## Affiliations

#### 1. Department of Chemistry, Manchester Institute of Biotechnology, The University of Manchester, Manchester, UK

Michael Winn, Michael Rowlinson, Fanghua Wang, Luis Bering, Daniel Francis, Colin Levy & Jason Micklefield

### Authors

#### 1. Michael Winn

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

#### 2. Michael Rowlinson

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

#### 3. Fanghua Wang

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

4. Luis Bering

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

5. Daniel Francis

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

6. Colin Levy

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

7. Jason Micklefield

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

## Contributions

M.W. and J.M. designed experiments; M.W., M.R., F.W., L.B. and D.F. carried out the experiments and provided additional experiment design; C.L. performed crystallographic studies. M.W. and J.M. wrote the manuscript. J.M. led the study.

## Corresponding author

Correspondence to [Jason Micklefield](#).

## Ethics declarations

## Competing interests

The authors declare no competing interests.

## Additional information

**Peer review information** *Nature* thanks Francesca Paradisi and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

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

## Extended data figures and tables

### Extended Data Fig. 1 Melting point determination of CfaL enzymes.

Melting point temperatures ( $T_m$ ) of the CfaLs in this study, obtained using a fluorescence-based assay conducted in a Bio-Rad CFX Connect qPCR machine. Higher  $T_m$  indicates improved thermal stability. The  $T_m$  is calculated as the lowest point when plotting the negative derivative of RFU (relative fluorescence units) as a function of temperature ( $dT$ ), versus the temperature (degrees Celsius).

### Extended Data Fig. 2 Rational mutagenesis of PbCfaL.

**a**, Structural comparison between PbCfaL (left) and the mutant PbCfaL(R395G) (right). R395 (circled) of PbCfaL (PDB ID 7A9I) is in the hinge region between the N-terminal domain (blue) and the flexible C-terminal domain (red). In PbCfaL(R395G) (PDB ID 7A9J) this large arginine residue is replaced by a much smaller glycine (circled) that is found in the other members of the CfaL family and many other similar ANL ligases. The overall structure of this mutant exhibits no other substantial structural difference from that of the wild type. **b**, Overlay of PbCfaL with three published ATP-dependent ligase structures (in ellipse) showing the conserved ATP binding location. When superimposed, PbCfaL (PDB ID 7A9I, blue), McbA (AMP bound, PDB ID 6SQ8, red), GrsA (ATP bound, PDB ID 1AMU, green) and AuaEII (anthranoyl-AMP bound, PDB ID 4WV3, light brown) show the conserved location of ATP binding. The corresponding loop in PbCfaL (inset, arrowed) is larger than in the other structures which may affect ATP binding. The location of this region within the structure of PbCfaL (grey) is also shown for reference. Structural alignment was performed using Chimera (version 1.14) MatchMaker.

### Extended Data Fig. 3 Synthesis and NMR spectra of crude product 64.

**a**, Preparative-scale synthesis of **64** from **10** and l-Ile catalysed by PbCfaL(R395G/A294P) (lysate), with either the addition of ATP (87% isolated yield) or recycling the endogenous ATP present in the lysate using the kinase (CHU) and polyphosphate (polyP) (52% isolated yield). **b**,  $^1\text{H}$ -NMR spectrum of crude product **64**. **c**,  $^{13}\text{C}$ -NMR spectrum of crude product **64**.

### Extended Data Fig. 4 Catalysed conversion of 10 and l-Ile to give 64.

**a**, Catalysed by PbCfaL (R395G/A294P) CLEAs. Reactions (100 mM Tris-HCl, 10 mM MgCl<sub>2</sub>, 10 mM ATP, 1 mM **10**, 3 mM l-isoleucine, 50 ml total volume) were run for 24 h; the CLEA (cross-linked enzyme aggregate) was then removed, washed, and reintroduced to an identical reaction. While activity was seen to reduce over the 5 days, the CLEA still retained high levels of productivity even after 5 recycles over 5 days, whereas cell lysates generally precipitated and lost all activity within 12 h. Although CfaL undergoes extensive conformational changes during catalysis, encapsulating it within CLEAs shows the potential of immobilization to extend the functional lifespan of the CfaL. More sophisticated immobilization techniques may have the potential to further retain activity. Conversion values were calculated from HPLC peak area ratios of product and starting materials, and represent means where  $n = 5$ , error bars denote s.d. **b**, Catalysed by purified PbCfaL(R395G/A294P), showing percentage conversions of **10** and l-Ile in the presence of various solvents and at different concentrations. Conversion values were calculated from HPLC peak area ratios of product and starting materials and represent means where  $n = 3$ , error bars denote s.d.

**Extended Data Fig. 5 LCMS analysis and extracted-ion chromatograms (EICs) of fragments of telaprevir and bortezomib synthesized by PbCfaL(R395G/A294P).**

**a**, Proposed route towards the antiviral agent telaprevir by CfaL (see reaction at top). The expected product of the reaction, **110**, was detected by LCMS (top trace, expected  $m/z$  262.1197, observed 262.1193 [M-H]<sup>-</sup>). Additional peaks consistent with dipeptide **111**, formed from condensation of two cyclohexylglycines, **58**, were also detected (bottom trace, **111a** and **111b**, expected  $m/z$  295.2027, observed 295.2036 [M-H]<sup>-</sup>). Although CfaL are highly selective for l-amino acid substrates, the appearance of two products of the same mass suggests formation of diastereomers, which may be due to a lack of enantioselectivity in the adenylation step forming the acyl donor when racemic cyclohexylglycine (**58**) is used. This indicates that **58** can function as both a carboxylic acid and an amine donor. **b**, Proposed route towards anti-cancer agent bortezomib via the synthesis of **112** by CfaL (see reaction at top). The expected product of the reaction was detected by LCMS (top trace, expected  $m/z$  270.0884, observed 270.0878 [M-H]<sup>-</sup>). An additional peak consistent with an l-Phe dipeptide (**113**) was also detected (bottom trace, expected  $m/z$  311.1401). This indicates that l-Phe can function as both acyl donor and amine acceptor. **c**, The reaction between carboxylic acid substrate (**9**), which is a good substrate for the enzyme, and cyclohexylglycine (**58**) gives only the desired product (**114**, top trace, expected  $m/z$  274.1449, observed 274.1460 [M-H]<sup>-</sup>). No cyclohexylglycine homodimer (dipeptide **111**) was evident in this case, indicating that homocoupling of **58** only takes place when carboxylic acid (acyl donor) substrates that are not well accepted by CfaL are used.

## **Extended Data Table 1 Table of kinetic parameters for adenylation reactions**

[Full size table](#)

## **Extended Data Table 2 Percentage yield of ligation reactions of carboxylic acids**

[Full size table](#)

## **Extended Data Table 3 Ligation of 9 to a range of proteinogenic amino acids**

[Full size table](#)

## **Extended Data Table 4 Ligation of 9 to a range of non-proteinogenic amino acids**

[Full size table](#)

# **Supplementary information**

## **Supplementary Information**

This file contains supplementary methods and compound characterization, NMR Spectra and supplementary references.

## **Reporting Summary**

## **Supplementary Figures**

This file contains supplementary text, supplementary figures 1 – 21, supplementary tables 1 – 4 and supplementary references.

# **Rights and permissions**

## **Reprints and Permissions**

# **About this article**



Check for  
updates

## **Cite this article**

Winn, M., Rowlinson, M., Wang, F. *et al.* Discovery, characterization and engineering of ligases for amide synthesis. *Nature* **593**, 391–398 (2021).

<https://doi.org/10.1038/s41586-021-03447-w>

## **Download citation**

- Received: 21 August 2020
- Accepted: 11 March 2021
- Published: 19 May 2021
- Issue Date: 20 May 2021
- DOI: <https://doi.org/10.1038/s41586-021-03447-w>

## Comments

By submitting a comment you agree to abide by our [Terms](#) and [Community Guidelines](#). If you find something abusive or that does not comply with our terms or guidelines please flag it as inappropriate.

[Download PDF](#)

Advertisement

---

This article was downloaded by **calibre** from <https://www.nature.com/articles/s41586-021-03447-w>

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

- Article
- [Published: 19 May 2021](#)

# Overwintering fires in boreal forests

- [Rebecca C. Scholten](#) ORCID: [orcid.org/0000-0002-0144-0572<sup>1</sup>](https://orcid.org/0000-0002-0144-0572),
- [Randi Jandt](#) ORCID: [orcid.org/0000-0001-8471-792X<sup>2</sup>](https://orcid.org/0000-0001-8471-792X),
- [Eric A. Miller](#) ORCID: [orcid.org/0000-0002-2021-2612<sup>3</sup>](https://orcid.org/0000-0002-2021-2612),
- [Brendan M. Rogers](#) ORCID: [orcid.org/0000-0001-6711-8466<sup>4</sup>](https://orcid.org/0000-0001-6711-8466) &
- [Sander Veraverbeke](#) ORCID: [orcid.org/0000-0003-1362-5125<sup>1</sup>](https://orcid.org/0000-0003-1362-5125)

[Nature](#) volume 593, pages 399–404 (2021) [Cite this article](#)

- 1546 Accesses
- 776 Altmetric
- [Metrics details](#)

## Subjects

- [Climate change](#)
- [Environmental sciences](#)
- [Fire ecology](#)
- [Natural hazards](#)

## Abstract

Forest fires are usually viewed within the context of a single fire season, in which weather conditions and fuel supply can combine to create conditions favourable for fire ignition—usually by lightning or human activity—and

spread<sup>1,2,3</sup>. But some fires exhibit ‘overwintering’ behaviour, in which they smoulder through the non-fire season and flare up in the subsequent spring<sup>4,5</sup>. In boreal (northern) forests, deep organic soils favourable for smouldering<sup>6</sup>, along with accelerated climate warming<sup>7</sup>, may present unusually favourable conditions for overwintering. However, the extent of overwintering in boreal forests and the underlying factors influencing this behaviour remain unclear. Here we show that overwintering fires in boreal forests are associated with hot summers generating large fire years and deep burning into organic soils, conditions that have become more frequent in our study areas in recent decades. Our results are based on an algorithm with which we detect overwintering fires in Alaska, USA, and the Northwest Territories, Canada, using field and remote sensing datasets. Between 2002 and 2018, overwintering fires were responsible for 0.8 per cent of the total burned area; however, in one year this amounted to 38 per cent. The spatiotemporal predictability of overwintering fires could be used by fire management agencies to facilitate early detection, which may result in reduced carbon emissions and firefighting costs.

## Access options

Subscribe to Journal

Get full journal access for 1 year

\$199.00

only \$3.90 per issue

[Subscribe](#)

All prices are NET prices.

VAT will be added later in the checkout.

Tax calculation will be finalised during checkout.

Rent or Buy article

Get time limited or full article access on ReadCube.

from \$8.99

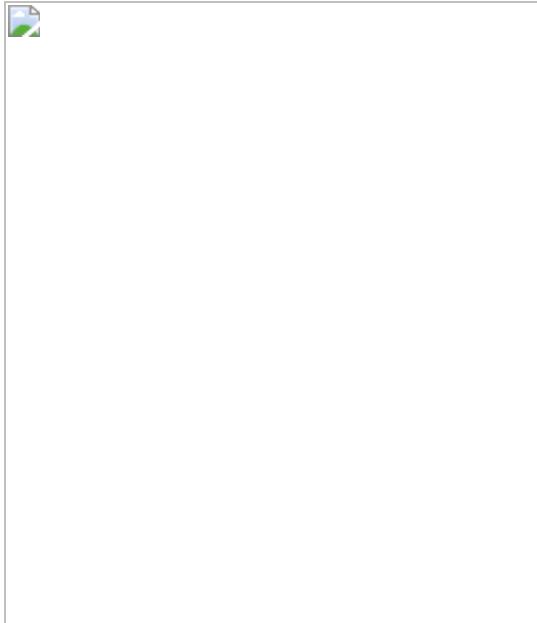
[Rent or Buy](#)

All prices are NET prices.

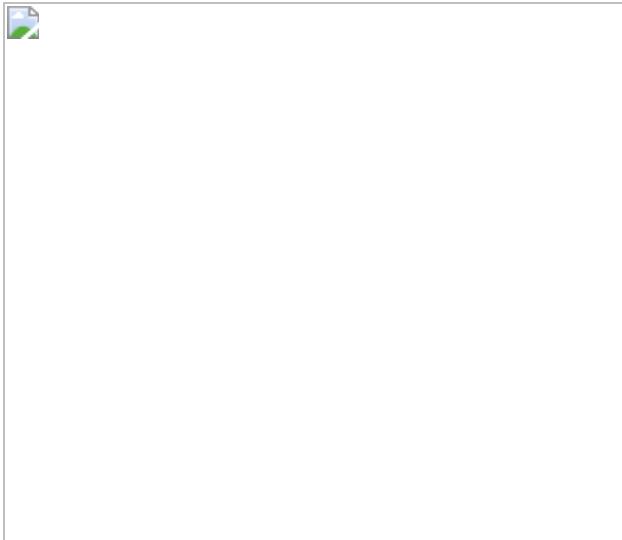
### **Additional access options:**

- [Log in](#)
- [Access through your institution](#)
- [Learn about institutional subscriptions](#)

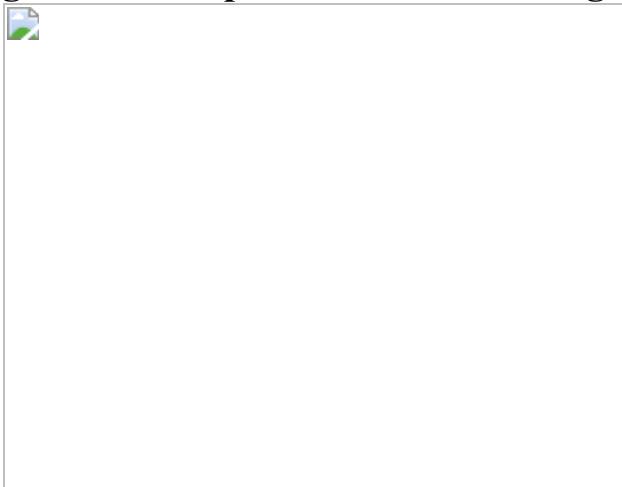
**Fig. 1: Landsat 8 false-colour time series of a 2015 fire in Alaska that generated an overwintering fire in 2016.**



**Fig. 2: Temporal drivers of overwintering fires and their long-term trends.**



**Fig. 3: Burn depth and overwintering.**



**Fig. 4: Occurrence of overwintering flare-ups.**

## **Data availability**

The location and timing of ignition of the overwintering fires used in this study are given in the [Supplementary Information](#). Daily burned area, emissions and ignitions data for Alaska and the Northwest Territories are archived at the Oak Ridge National Laboratory Distributed Active Archive Center for biogeochemical dynamics (<https://doi.org/10.3334/ORNLDAAC/1812>). Lightning data are available from the Alaska Interagency Coordination Center (<https://fire.ak.blm.gov/predsvcs/maps.php>) and from Environment and Climate Change Canada. Infrastructure data are available for Alaska from the Department of Natural Resources (<https://catalog.data.gov/dataset/alaska-infrastructure-1-63360>), and for the Northwest Territories from Statistics Canada (<https://www150.statcan.gc.ca/n1/en/catalogue/92-500-X>) and the Government of Yukon ([https://hub.arcgis.com/datasets/322b6cf3fa1444c289a1d611a4778ead\\_42/ata](https://hub.arcgis.com/datasets/322b6cf3fa1444c289a1d611a4778ead_42/ata)). MODSCAG snow fraction data are freely available from the JPL Snow Data Server (<http://snow.jpl.nasa.gov/portal/>). All climate data used in this study are available from the North America Regional Reanalysis (<https://psl.noaa.gov/data/gridded/data.narr.html>). All data used for the analysis of spatial drivers are freely available, including the ArcticDEM (<https://doi.org/10.7910/DVN/OHHUKH>), the Northern Circumpolar Soil Carbon Database v2 (<https://doi.org/10.5879/ECDS/00000002>) and the Fuel Characteristic Classification System (<https://www.landfire.gov/fccs.php>).

## Code availability

Codes used to analyse the data are available from <https://github.com/screbec/Overwintering-fires> or <https://doi.org/10.5281/zenodo.4549321>.

## References

1. 1.

Sedano, F. & Randerson, J. T. Multi-scale influence of vapor pressure deficit on fire ignition and spread in boreal forest ecosystems.

*Biogeosciences* **11**, 3739–3755 (2014).

[ADS](#) [Google Scholar](#)

2. 2.

Veraverbeke, S. et al. Lightning as a major driver of recent large fire years in North American boreal forests. *Nat. Clim. Chang.* **7**, 529–534 (2017).

[ADS](#) [Google Scholar](#)

3. 3.

Calef, M. P., McGuire, A. D. & Chapin, F. S. Human influences on wildfire in Alaska from 1988 through 2005: an analysis of the spatial patterns of human impacts. *Earth Interact.* **12**, 1–17 (2008).

[ADS](#) [Google Scholar](#)

4. 4.

McCarty, J. L., Smith, T. E. L. & Turetsky, M. R. Arctic fires re-emerging. *Nat. Geosci.* **13**, 658–660 (2020).

[ADS](#) [CAS](#) [Google Scholar](#)

5. 5.

Irannezhad, M., Liu, J., Ahmadi, B. & Chen, D. The dangers of Arctic zombie wildfires. *Science* **369**, 1171 (2020).

[ADS](#) [Google Scholar](#)

6. 6.

Rein, G. in *Fire Phenomena and the Earth System: An Interdisciplinary Guide to Fire Science* (ed. Belcher, C. M.) 15–34 (Wiley-Blackwell, 2013).

7. 7.

Post, E. et al. The polar regions in a 2 °C warmer world. *Sci. Adv.* **5**, eaaw9883 (2019).

[ADS](#) [CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

8. 8.

Overland, J. E., Wang, M., Walsh, J. E. & Stroeve, J. C. Future Arctic climate changes: adaptation and mitigation time scales. *Earth's Future* **2**, 68–74 (2014).

[ADS](#) [Google Scholar](#)

9. 9.

Tarnocai, C. et al. Soil organic carbon pools in the northern circumpolar permafrost region. *Glob. Biogeochem. Cycles* **23**, GB2023 (2009).

[ADS](#) [Google Scholar](#)

10. 10.

Walker, X. J. et al. Increasing wildfires threaten historic carbon sink of boreal forest soils. *Nature* **572**, 520–523 (2019).

[ADS](#) [CAS](#) [Google Scholar](#)

11. 11.

Turetsky, M. R. et al. Recent acceleration of biomass burning and carbon losses in Alaskan forests and peatlands. *Nat. Geosci.* **4**, 27–31 (2011).

[ADS](#) [CAS](#) [Google Scholar](#)

12. 12.

Walker, X. J. et al. Soil organic layer combustion in boreal black spruce and jack pine stands of the Northwest Territories, Canada. *Int. J. Wildl. Fire* **27**, 125–134 (2018).

[Google Scholar](#)

13. 13.

Turetsky, M. R. et al. Global vulnerability of peatlands to fire and carbon loss. *Nat. Geosci.* **8**, 11–14 (2015).

[ADS](#) [CAS](#) [Google Scholar](#)

14. 14.

Flannigan, M. D. et al. Fuel moisture sensitivity to temperature and precipitation: climate change implications. *Clim. Change* **134**, 59–71 (2016).

[ADS](#) [CAS](#) [Google Scholar](#)

15. 15.

Coops, N. C., Hermosilla, T., Wulder, M. A., White, J. C. & Bolton, D. K. A thirty year, fine-scale, characterization of area burned in Canadian forests shows evidence of regionally increasing trends in the last decade. *PLoS One* **13**, e0197218 (2018).

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

16. 16.

USDA Forest Service, USFS-USDI and NASF. *Large Fire Cost Reduction Action Plan*  
[https://www.fs.usda.gov/sites/default/files/media\\_wysiwyg/5100\\_large\\_firecostreductionaction\\_mar\\_03.pdf](https://www.fs.usda.gov/sites/default/files/media_wysiwyg/5100_large_firecostreductionaction_mar_03.pdf) (2003).

17. 17.

Podur, J. & Wotton, M. Will climate change overwhelm fire management capacity? *Ecol. Modell.* **221**, 1301–1309 (2010).

[Google Scholar](#)

18. 18.

Tymstra, C., Stocks, B. J., Cai, X. & Flannigan, M. D. Wildfire management in Canada: review, challenges and opportunities. *Prog. Disaster Sci.* **5**, 100045 (2020); erratum **8**, 100045 (2020).

[Google Scholar](#)

19. 19.

Stocks, B. J. et al. Large forest fires in Canada, 1959–1997. *J. Geophys. Res.* **107**, <https://doi.org/10.1029/2001JD000484> (2002).

20. 20.

Wiggins, E. B. et al. Evidence for a larger contribution of smoldering combustion to boreal forest fire emissions from tower observations in Alaska. *Atmos. Chem. Phys.* <https://doi.org/10.5194/acp-2019-1067> (in the press).

21. 21.

Rein, G., Garcia, J., Simeoni, A., Tihay, V. & Ferrat, L. Smouldering natural fires: comparison of burning dynamics in boreal peat and Mediterranean humus. *WIT Trans. Ecol. Environ.* **119**, 183–192 (2008).

[Google Scholar](#)

22. 22.

Baber, C. & McMaster, R. 2019 Alaska Statewide Annual Operating Plan. <https://fire.ak.blm.gov/administration/asma.php> (Alaska Statewide Master Agreement, 2019).

23. 23.

Alaska Interagency Coordination Center. 2010 Alaska fire statistics. <https://www.frames.gov/catalog/12055> (Wildland Fire Summary and Statistics Annual Report, 2010).

24. 24.

Alaska Division of Forestry. State Forestry monitoring hot spots that overwintered from Deshka Landing Fire. <https://akfireinfo.com/2020/04/10/state-forestry-monitoring-hot-spots-that-overwintered-from-deshka-landing-fire/> (2020).

25. 25.

Giglio, L., Schroeder, W. & Justice, C. O. The collection 6 MODIS active fire detection algorithm and fire products. *Remote Sens. Environ.* **178**, 31–41 (2016).

[ADS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

26. 26.

Kasischke, E. S., Rupp, T. S. & Verbyla, D. L. in *Alaska's Changing Boreal Forest* (eds Chapin, F. S. III, Oswood, M. et al.) 285–301 (Oxford Univ. Press, 2006).

27. 27.

Westerling, A. L., Hidalgo, H. G., Cayan, D. R. & Swetnam, T. W. Warming and earlier spring increase western U.S. forest wildfire activity. *Science* **313**, 940–943 (2006).

[ADS](#) [CAS](#) [Google Scholar](#)

28. 28.

Painter, T. H. et al. Retrieval of subpixel snow covered area, grain size, and albedo from MODIS. *Remote Sens. Environ.* **113**, 868–879 (2009).

[ADS](#) [Google Scholar](#)

29. 29.

Scholten, R. C., Jandt, R. R., Miller, E. A., Rogers, B. M. & Veraverbeke, S. ABoVE: Ignitions, burned area and emissions of fires in AK, YT, and NWT, 2001–2018.  
<https://doi.org/10.3334/ORNLDAA/1812> (2020).

30. 30.

Xiao, J. & Zhuang, Q. Drought effects on large fire activity in Canadian and Alaskan forests. *Environ. Res. Lett.* **2**, 044003 (2007).

[ADS](#) [Google Scholar](#)

31. 31.

Flannigan, M. D. et al. Global wildland fire season severity in the 21st century. *For. Ecol. Manage.* **294**, 54–61 (2013).

[Google Scholar](#)

32. 32.

Jolly, W. M. et al. Climate-induced variations in global wildfire danger from 1979 to 2013. *Nat. Commun.* **6**, 7537 (2015).

[ADS](#) [CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

33. 33.

Adams, W. H. *The Role of Fire in the Alaska Taiga. An Unsolved Problem* (Bureau of Land Management, State Office, Anchorage, AK, 1974); preprint at <https://scholarworks.alaska.edu/handle/11122/6675> (2016).

34. 34.

Certini, G. Effects of fire on properties of forest soils: a review. *Oecologia* **143**, 1–10 (2005).

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

35. 35.

Kane, E. S., Kasischke, E. S., Valentine, D. W., Turetsky, M. R. & McGuire, A. D. Topographic influences on wildfire consumption of soil organic carbon in interior Alaska: implications for black carbon accumulation. *J. Geophys. Res. Biogeosci.* **112**, 1–11 (2007).

[Google Scholar](#)

36. 36.

Hoy, E. E., Turetsky, M. R. & Kasischke, E. S. More frequent burning increases vulnerability of Alaskan boreal black spruce forests. *Environ. Res. Lett.* **11**, 095001 (2016).

[ADS](#) [Google Scholar](#)

37. 37.

Miyanishi, K. & Johnson, E. A. Process and patterns of duff consumption in the mixedwood boreal forest. *Can. J. For. Res.* **32**, 1285–1295 (2002).

[Google Scholar](#)

38. 38.

Kasischke, E. S. & Turetsky, M. R. Recent changes in the fire regime across the North American boreal region — spatial and temporal patterns of burning across Canada and Alaska. *Geophys. Res. Lett.* **33**, <https://doi.org/10.1029/2006GL025677> (2006).

39. 39.

Johnstone, J. F. et al. Factors shaping alternate successional trajectories in burned black spruce forests of Alaska. *Ecosphere* **11**, <https://doi.org/10.1002/ecs2.3129> (2020).

40. 40.

Mekonnen, Z. A., Riley, W. J., Randerson, J. T., Grant, R. F. & Rogers, B. M. Expansion of high-latitude deciduous forests driven by interactions between climate warming and fire. *Nat. Plants* **5**, 952–958 (2019).

[Google Scholar](#)

41. 41.

Andreae, M. O. & Merlet, P. Emission of trace gases and aerosols from biomass burning. *Glob. Biogeochem. Cycles* **15**, 955–966 (2001).

[ADS](#) [CAS](#) [Google Scholar](#)

42. 42.

Dean, J. F. et al. Methane feedbacks to the global climate system in a warmer world. *Rev. Geophys.* **56**, 207–250 (2018).

[ADS](#) [Google Scholar](#)

43. 43.

Beaudoin, A., Bernier, P. Y., Villemaire, P., Guindon, L. & Guo, X. J. Tracking forest attributes across Canada between 2001 and 2011 using a k nearest neighbors mapping approach applied to MODIS imagery. *Can. J. For. Res.* **48**, 85–93 (2018).

[Google Scholar](#)

44. 44.

Veraverbeke, S., Rogers, B. M. & Randerson, J. T. Daily burned area and carbon emissions from boreal fires in Alaska. *Biogeosci. Discuss.* **12**, 3579–3601 (2015).

[ADS](#) [CAS](#) [Google Scholar](#)

45. 45.

Kasischke, E. S. et al. Quantifying burned area for North American forests: implications for direct reduction of carbon stocks. *J. Geophys. Res. Biogeosci.* **116**, 1–17 (2011).

[Google Scholar](#)

46. 46.

Farukh, M. A. & Hayasaka, H. Active forest fire occurrences in severe lightning years in Alaska. *J. Nat. Disaster Sci.* **33**, 71–84 (2012).

[Google Scholar](#)

47. 47.

Burrows, W. R. & Kochtubajda, B. A decade of cloud-to-ground lightning in Canada: 1999-2008. Part 1: flash density and occurrence. *Atmos.-Ocean* **48**, 177–194 (2010).

[Google Scholar](#)

48. 48.

Bieniek, P. A. et al. Lightning variability in dynamically downscaled simulations of Alaska's present and future summer climate. *J. Appl. Meteorol. Climatol.* **59**, 1139–1152 (2020).

[ADS](#) [Google Scholar](#)

49. 49.

Kochtubajda, B. et al. Exceptional cloud-to-ground lightning during an unusually warm summer in Yukon, Canada. *J. Geophys. Res. Atmos.* **116**, <https://doi.org/10.1029/2011JD016080> (2011).

50. 50.

Kochtubajda, B., Stewart, R. & Tropea, B. Lightning and weather associated with the extreme 2014 wildfire season in Canada's Northwest Territories. In *Proceedings of the 24th International Lightning Detection Conference* 1–4 (VAISALA, 2016).

51. 51.

Dowdy, A. J. & Mills, G. A. Atmospheric and fuel moisture characteristics associated with lightning-attributed fires. *J. Appl. Meteorol. Climatol.* **51**, 2025–2037 (2012).

[ADS](#) [Google Scholar](#)

52. 52.

Larjavaara, M., Pennanen, J. & Tuomi, T. J. Lightning that ignites forest fires in Finland. *Agric. For. Meteorol.* **132**, 171–180 (2005).

[ADS](#) [Google Scholar](#)

53. 53.

Duncan, B. W., Adrian, F. W. & Stolen, E. D. Isolating the lightning ignition regime from a contemporary background fire regime in east-central Florida, USA. *Can. J. For. Res.* **40**, 286–297 (2010).

[Google Scholar](#)

54. 54.

Veraverbeke, S. et al. Mapping the daily progression of large wildland fires using MODIS active fire data. *Int. J. Wildl. Fire* **23**, 655–667 (2014).

[Google Scholar](#)

55. 55.

Statistics Canada. Road Network File 2010.

<https://www150.statcan.gc.ca/n1/en/catalogue/92-500-X> (2016).

56. 56.

Government of Yukon. Corporate Spatial Warehouse.

[ftp://ftp.geomaticsyukon.ca/GeoYukon/Transportation/Roads\\_1M/](ftp://ftp.geomaticsyukon.ca/GeoYukon/Transportation/Roads_1M/) (2018).

57. 57.

Rittger, K., Painter, T. H. & Dozier, J. Assessment of methods for mapping snow cover from MODIS. *Adv. Water Resour.* **51**, 367–380 (2013).

[ADS](#) [Google Scholar](#)

58. 58.

Gallant, A. L., Binnian, E. F., Omernik, J. M. & Shasby, M. B. *Ecoregions of Alaska* (Professional Paper 1567, USGS, 1995).

59. 59.

Canadian Council on Ecological Areas (CCEA). Canada ecozones.  
<https://cceaa-ccae.org/ecozone-downloads/> (2016).

60. 60.

Mesinger, F. et al. North American regional reanalysis. *Bull. Am. Meteorol. Soc.* **87**, 343–360 (2006).

[ADS](#) [Google Scholar](#)

61. 61.

Van Wagner, C. E. *Development and Structure of the Canadian Fire Weather Index System*. Forestry Technical Report Vol. 35 (Canadian Forestry Service Headquarters, Ottawa, 1987).

62. 62.

York, A. D. & Jandt, R. R. *Opportunities to Apply Remote Sensing in Boreal/Arctic Wildfire Management & Science: A Workshop Report* [www.frames.gov/catalog/57849](http://www.frames.gov/catalog/57849) (University of Alaska, Fairbanks, 2019).

63. 63.

Schroeder, W., Oliva, P., Giglio, L. & Csizar, I. A. The New VIIRS 375m active fire detection data product: algorithm description and initial assessment. *Remote Sens. Environ.* **143**, 85–96 (2014).

[ADS](#) [Google Scholar](#)

64. 64.

Welch, B. L. The significance of the difference between two means when the population variances are unequal. *Biometrika* **29**, 350–362 (1938).

[MATH](#) [Google Scholar](#)

65. 65.

Welch, B. L. The generalization of ‘Student’s’ problem when several different population variances are involved. *Biometrika* **34**, 28–35 (1947).

[MathSciNet](#) [CAS](#) [MATH](#) [Google Scholar](#)

66. 66.

Morin, P. et al. ArcticDEM; a publically available, high resolution elevation model of the Arctic. *Geophys. Res. Abstr.* **18**, EGU2016-

8396 (2016).

[Google Scholar](#)

67. 67.

Porter, C. et al. ArcticDEM. <https://doi.org/10.7910/DVN/OHHUKH> (Harvard Dataverse, 2018).

68. 68.

Dai, C., Durand, M., Howat, I. M., Altenau, E. H. & Pavelsky, T. M. Estimating river surface elevation from arcticDEM. *Geophys. Res. Lett.* **45**, 3107–3114 (2018).

[ADS](#) [Google Scholar](#)

69. 69.

Hansen, M. C. et al. Global percent tree cover at a spatial resolution of 500 meters: first results of the MODIS vegetation continuous fields algorithm. *Earth Interact.* **7**, 1–15 (2003).

[Google Scholar](#)

70. 70.

Pettinari, M. L. & Chuvieco, E. Generation of a global fuel data set using the fuel characteristic classification system. *Biogeosciences* **13**, 2061–2076 (2016).

[ADS](#) [Google Scholar](#)

71. 71.

Ottmar, R. D., Sandberg, D. V., Riccardi, C. L. & Prichard, S. J. An overview of the fuel characteristic classification system — quantifying, classifying, and creating fuelbeds for resource planning. *Can. J. For. Res.* **37**, 2383–2393 (2007).

[Google Scholar](#)

72. 72.

Riccardi, C. L. et al. The fuelbed: a key element of the fuel characteristic classification system. *Can. J. For. Res.* **37**, 2394–2412 (2007).

[Google Scholar](#)

73. 73.

Beaudoin, A., Bernier, P. Y., Villemaire, P., Guindon, L. & Guo, X. *Species Composition, Forest Properties and Land Cover Types Across Canada's Forests at 250m Resolution for 2001 and 2011*. <https://doi.org/10.23687/ec9e2659-1c29-4ddb-87a2-6aced147a990> (Natural Resources Canada, Canadian Forest Service, Laurentian Forest Centre, 2017).

74. 74.

Hugelius, G. et al. The northern circumpolar soil carbon database: spatially distributed datasets of soil coverage and soil carbon storage in the northern permafrost regions. *Earth Syst. Sci. Data* **5**, 3–13 (2013).

[ADS](#) [Google Scholar](#)

[Download references](#)

## Acknowledgements

We would like to thank C. Van Der Horn and G. Branson (Alaska Interagency Coordination Center), and M. Coyle (Forest Management Division, Department of Environment and Natural Resources, Government of the Northwest Territories), for providing ground truth data on overwintering fires. We wish to thank Environment and Climate Change Canada for their generous permission to use Canadian Lightning Detection Network data, and the Bureau of Land Management, Alaska Fire Service,

for providing cost information. We thank NASA JPL's Snow Data Center for making their MODSCAG data available. This work was funded by the Netherlands Organization for Scientific Research (NWO) through a Vidi grant (Fires Pushing Trees North) awarded to S.V. B.M.R. acknowledges support from the National Aeronautics and Space Administration (NASA) Arctic-Boreal Vulnerability Experiment (ABOVE; NNX15AU56A).

## Author information

### Affiliations

1. Faculty of Science, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands  
Rebecca C. Scholten & Sander Veraverbeke
2. Alaska Fire Science Consortium, University of Alaska, Fairbanks, AK, USA  
Randi Jandt
3. Bureau of Land Management, Alaska Fire Service, Fort Wainwright, AK, USA  
Eric A. Miller
4. Woodwell Climate Research Center, Falmouth, MA, USA  
Brendan M. Rogers

### Authors

1. Rebecca C. Scholten  
[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

2. Randi Jandt

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

3. Eric A. Miller

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

4. Brendan M. Rogers

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

5. Sander Veraverbeke

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

## Contributions

S.V. and R.C.S. designed the research; R.C.S. performed the analysis with input from S.V.; B.M.R. contributed to the interpretation of cost data; R.R.J. and E.A.M. contributed to the interpretation of field data; R.C.S. drafted the manuscript; and all authors participated in manuscript editing.

## Corresponding author

Correspondence to [Rebecca C. Scholten](#).

## Ethics declarations

## Competing interests

The authors declare no competing interests.

## Additional information

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

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

## Extended data figures and tables

### [Extended Data Fig. 1 Aerial view of the Seven Mile Slough fire in Alaska on 9 May 2011.](#)

Smouldering hotspots (labelled with 'a') had overwintered and burned in the organic soil layer below the spruces of an unburned island. Green tree crowns of the fallen trees (labelled 'b') in the original unburned island (perimeter approximated in black) suggest that tree roots were damaged by subsurface burning. (Photograph by E.A.M.).

### [Extended Data Fig. 2 Workflow used to detect large overwintering fires.](#)

First, ignition locations, dates and causes according to official fire databases were extracted. In four steps, the algorithm filters these ignitions by date, distance to an old burn scar, and co-occurrence of lightning strikes and infrastructure elements. Small overwintering fires that were not detected by satellite products were used to derive thresholds for the algorithm.

### [Extended Data Fig. 3 Spatiotemporal characteristics of small overwintering fires.](#)

Overwintering fires emerge earlier after the seasonal snow melt (**a**) and closer to a fire scar from the year before (**b**) than other fires. 'Other fires' refer to all fires not classified as overwintering in official fire databases. Day since regional snow melt was taken from government sources.

### [Extended Data Fig. 4 Lag times and distance to roads.](#)

Shown are histograms of lag time between lightning strikes and ignition detections (**a**), and distance to road for ignitions by humans (**b**). Human and lightning ignitions were characterized on the basis of the Alaskan Wildland Fire Maps (Alaska, AK) and the Canadian National Fire Database (Northwest Territories, NWT). The black lines indicate the thresholds used to eliminate potential overwintering fires due to spatial proximity to infrastructure and spatiotemporal proximity to lightning strikes.

**Extended Data Fig. 5 Average and extreme temperature trends for interior Alaska and the Northwest Territories.**

**a, b**, Average of the daily maximum temperature of the summer months May–September; **c, d**, its 90th percentile. **e, f**, Number of hot days surpassing the 90th percentile. Panels **a, c, e** show data for interior Alaska, and panels **b, d, f** for the taiga plains and the taiga shield of the Northwest Territories.

**Extended Data Fig. 6 Scatter plots and Spearman correlations ( $\rho$ ) of summer temperature, burned area and overwintering flare-ups.**

**a, b**, Daily mean maximum temperature between May and September (MJJAS;  $T_{\text{mean}}$ ) and annual burned area. **c, d**, Previous year's burned area and the number of overwintering flare-ups. **e, f**, MJJAS maximum temperature and the number of overwintering flare-ups. Panels **a, c, e** show data for Alaska, and panels **b, d, f** for the Northwest Territories.

**Extended Data Fig. 7 Scatter plots and Spearman correlations of temperature extremes and burned area, overwintering flare-ups and burn depth.**

**a, b**, Number of MJJAS hot days (days with a maximum temperature hotter than the 1979–2020 90th percentile,  $T_{90}$ ) and burned area. **c, d**, Number of MJJAS hot days and overwintering ignitions. **e, f**, 90th percentile of MJJAS

temperature ( $T_{\max 90}$ ) and average burn depth. Panels **a**, **c**, **e** show data for Alaska, and panels **b**, **d**, **f** for the Northwest Territories.

**Extended Data Table 1 Correlation of meteorology and the number of overwintering flare-ups**

[Full size table](#)

**Extended Data Table 2 Comparison of burn scars with and without overwintering fires**

[Full size table](#)

**Extended Data Table 3 Effect of spring meteorology on the size of overwintering fires**

[Full size table](#)

## Supplementary information

### [Supplementary Information](#)

This file contains Supplementary Table 1 and Supplementary Table 2.

### [Peer Review File](#)

## Rights and permissions

### [Reprints and Permissions](#)

## About this article



Check for  
updates

## Cite this article

Scholten, R.C., Jandt, R., Miller, E.A. *et al.* Overwintering fires in boreal forests. *Nature* **593**, 399–404 (2021). <https://doi.org/10.1038/s41586-021->

03437-y

[Download citation](#)

- Received: 16 July 2020
- Accepted: 09 March 2021
- Published: 19 May 2021
- Issue Date: 20 May 2021
- DOI: <https://doi.org/10.1038/s41586-021-03437-y>

## Comments

By submitting a comment you agree to abide by our [Terms](#) and [Community Guidelines](#). If you find something abusive or that does not comply with our terms or guidelines please flag it as inappropriate.

[Access through your institution](#)

[Change institution](#)

[Buy or subscribe](#)

Advertisement

---

This article was downloaded by **calibre** from <https://www.nature.com/articles/s41586-021-03437-y>.

- Article
- [Published: 28 April 2021](#)

# Somatic mutation landscapes at single-molecule resolution

- [Federico Abascal](#) [ORCID: orcid.org/0000-0002-6201-1587<sup>1</sup>](#),
- [Luke M. R. Harvey<sup>1</sup>](#) [na1](#),
- [Emily Mitchell<sup>1,2</sup>](#) [na1](#),
- [Andrew R. J. Lawson](#) [ORCID: orcid.org/0000-0003-3592-1005<sup>1</sup>](#) [na1](#),
- [Stefanie V. Lensing<sup>1</sup>](#) [na1](#),
- [Peter Ellis<sup>1</sup>](#) [na1](#) [nAff9](#),
- [Andrew J. C. Russell](#) [ORCID: orcid.org/0000-0001-5411-2807<sup>1</sup>](#),
- [Raul E. Alcantara<sup>1</sup>](#),
- [Adrian Baez-Ortega<sup>1</sup>](#),
- [Yichen Wang](#) [ORCID: orcid.org/0000-0001-5122-7545<sup>1</sup>](#),
- [Eugene Jing Kwa<sup>1</sup>](#),
- [Henry Lee-Six](#) [ORCID: orcid.org/0000-0003-4831-8088<sup>1</sup>](#),
- [Alex Cagan](#) [ORCID: orcid.org/0000-0002-7857-4771<sup>1</sup>](#),
- [Tim H. H. Coorens](#) [ORCID: orcid.org/0000-0002-5826-3554<sup>1</sup>](#),
- [Michael Spencer Chapman<sup>1</sup>](#),
- [Sigurgeir Olafsson<sup>1</sup>](#),
- [Steven Leonard<sup>1</sup>](#),
- [David Jones](#) [ORCID: orcid.org/0000-0002-0407-0386<sup>1</sup>](#),
- [Heather E. Machado<sup>1</sup>](#),
- [Megan Davies<sup>2</sup>](#),
- [Nina F. Øbro<sup>2,3</sup>](#),
- [Krishnaa T. Mahubani](#) [ORCID: orcid.org/0000-0002-1327-2334<sup>3,4,5</sup>](#),
- [Kieren Allinson<sup>6</sup>](#),
- [Moritz Gerstung](#) [ORCID: orcid.org/0000-0001-6709-963X<sup>7</sup>](#),

- [Kurosh Saeb-Parsy](#) [ORCID: orcid.org/0000-0002-0633-3696<sup>4,5</sup>](#),
- [David G. Kent<sup>2,8</sup>](#),
- [Elisa Laurenti](#) [ORCID: orcid.org/0000-0002-9917-9092<sup>2,3</sup>](#),
- [Michael R. Stratton](#) [ORCID: orcid.org/0000-0001-6035-153X<sup>1</sup>](#),
- [Raheleh Rahbari](#) [ORCID: orcid.org/0000-0002-1839-7785<sup>1</sup>](#),
- [Peter J. Campbell](#) [ORCID: orcid.org/0000-0002-3921-0510<sup>1,3</sup>](#),
- [Robert J. Osborne](#) [ORCID: orcid.org/0000-0002-1914-1239<sup>1</sup>](#) nAff10  
&
- [Iñigo Martincorena](#) [ORCID: orcid.org/0000-0003-1122-4416<sup>1</sup>](#)

[Nature](#) volume **593**, pages 405–410 (2021) [Cite this article](#)

- 18k Accesses
- 526 Altmetric
- [Metrics details](#)

## Subjects

- [Cancer genomics](#)
- [DNA sequencing](#)
- [Genomics](#)
- [Mutation](#)

## Abstract

Somatic mutations drive the development of cancer and may contribute to ageing and other diseases<sup>1,2</sup>. Despite their importance, the difficulty of detecting mutations that are only present in single cells or small clones has limited our knowledge of somatic mutagenesis to a minority of tissues. Here, to overcome these limitations, we developed nanorate sequencing (NanoSeq), a duplex sequencing protocol with error rates of less than five errors per billion base pairs in single DNA molecules from cell populations. This rate is two orders of magnitude lower than typical somatic mutation loads, enabling the study of somatic mutations in any tissue independently

of clonality. We used this single-molecule sensitivity to study somatic mutations in non-dividing cells across several tissues, comparing stem cells to differentiated cells and studying mutagenesis in the absence of cell division. Differentiated cells in blood and colon displayed remarkably similar mutation loads and signatures to their corresponding stem cells, despite mature blood cells having undergone considerably more divisions. We then characterized the mutational landscape of post-mitotic neurons and polyclonal smooth muscle, confirming that neurons accumulate somatic mutations at a constant rate throughout life without cell division, with similar rates to mitotically active tissues. Together, our results suggest that mutational processes that are independent of cell division are important contributors to somatic mutagenesis. We anticipate that the ability to reliably detect mutations in single DNA molecules could transform our understanding of somatic mutagenesis and enable non-invasive studies on large-scale cohorts.

## Access options

Subscribe to Journal

Get full journal access for 1 year

\$199.00

only \$3.90 per issue

[Subscribe](#)

All prices are NET prices.

VAT will be added later in the checkout.

Tax calculation will be finalised during checkout.

Rent or Buy article

Get time limited or full article access on ReadCube.

from \$8.99

[Rent or Buy](#)

All prices are NET prices.

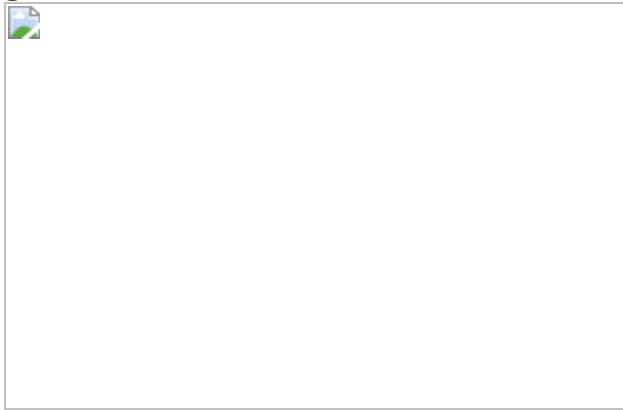
### **Additional access options:**

- [Log in](#)
- [Access through your institution](#)
- [Learn about institutional subscriptions](#)

**Fig. 1: BotSeqS and NanoSeq sequencing protocols.**

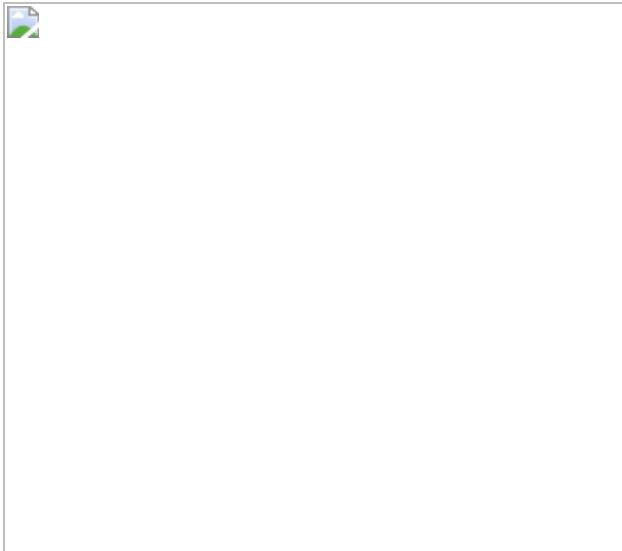


**Fig. 2: Mutation in stem and differentiated cells.**



**Fig. 3: Mutational landscape in neurons and smooth muscle.**





## Data availability

Information on data availability for all samples is available in Supplementary Table 1. NanoSeq sequencing data have been deposited in the European Genome-phenome Archive (EGA; <https://www.ebi.ac.uk/ega/>) under accession number EGAD00001006459. Sperm samples are available from the EGA under accession number EGAD00001007028. Standard sequencing data have been deposited in the EGA under accession number EGAD00001006595. For publicly available samples, references to the original sources are provided in Supplementary Table 1. Substitution and indel rates are available in Supplementary Table 4. Substitution and indel calls for samples sequenced with NanoSeq are available in Supplementary Tables 5, 6. Trinucleotide substitution profiles are available in Supplementary Table 7. A detailed NanoSeq protocol is available in Protocol Exchange<sup>53</sup>.

## Code availability

The bioinformatics pipeline to process NanoSeq sequencing data comprises all steps including processing sequencing data, mapping, calling mutations and calculating corrected burden estimates and substitution profiles. This code is available from <https://zenodo.org/record/4604537> (<https://doi.org/10.5281/zenodo.4604537>). Pipelines to call indels, perform

signature extraction and signature fitting with sigfit, simulate the efficiency of the NanoSeq protocol, calculate the mutation burden in specific genomic regions and reproduce most of the main plots are also available from <https://zenodo.org/record/4604537>. Analyses in R were done with R v.3.3 and v.3.6. R libraries used include: GenomicRanges<sup>54</sup> (v.1.38.0), Rsamtools (v.2.2.3), MASS (v.7.3-51.5), sigfit<sup>52</sup> (v.2.0), readxl (v.1.3.1), deconstructSigs (v.1.8.0), lsa (v.0.73.2), deepSNV<sup>55</sup> (v.1.32.0), lme4 (v.1.1-26), afex (v.0.28-1), lmerTest (v.3.1-3), bootpredictlme4 (v.0.1) and Biostrings (v.2.54.0). Our pipeline makes use of samtools<sup>56</sup> v.1.9, bcftools<sup>57</sup> v.1.9, bwa v.0.7.5a-r405 and bedtools<sup>58</sup> v.2.29.0. We also used the following software: CaVeMan (v.2020), Pindel (v.2020) and MPBoot v.1.1.0.

## References

1. 1.

Kennedy, S. R., Loeb, L. A. & Herr, A. J. Somatic mutations in aging, cancer and neurodegeneration. *Mech. Ageing Dev.* **133**, 118–126 (2012).

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

2. 2.

Vogelstein, B. et al. Cancer genome landscapes. *Science* **339**, 1546–1558 (2013).

[ADS](#) [CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

3. 3.

Martincorena, I. et al. High burden and pervasive positive selection of somatic mutations in normal human skin. *Science* **348**, 880–886 (2015)

[ADS](#) [CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

4. 4.

Martincorena, I. et al. Somatic mutant clones colonize the human esophagus with age. *Science* **362**, 911–917 (2018).

[ADS](#) [CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

5. 5.

Yizhak, K. et al. RNA sequence analysis reveals macroscopic somatic clonal expansion across normal tissues. *Science* **364**, eaaw0726 (2019).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

6. 6.

Lee-Six, H. et al. The landscape of somatic mutation in normal colorectal epithelial cells. *Nature* **574**, 532–537 (2019).

[ADS](#) [CAS](#) [PubMed](#) [Google Scholar](#)

7. 7.

Brunner, S. F. et al. Somatic mutations and clonal dynamics in healthy and cirrhotic human liver. *Nature* **574**, 538–542 (2019).

[ADS](#) [CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

8. 8.

Li, R. et al. Macroscopic somatic clonal expansion in morphologically normal human urothelium. *Science* **370**, 82–89 (2020).

[ADS](#) [CAS](#) [PubMed](#) [Google Scholar](#)

9. 9.

Welch, J. S. et al. The origin and evolution of mutations in acute myeloid leukemia. *Cell* **150**, 264–278 (2012).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

10. 10.

Blokzijl, F. et al. Tissue-specific mutation accumulation in human adult stem cells during life. *Nature* **538**, 260–264 (2016).

[ADS](#) [CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

11. 11.

Franco, I. et al. Somatic mutagenesis in satellite cells associates with human skeletal muscle aging. *Nat. Commun.* **9**, 800 (2018).

[ADS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

12. 12.

Lodato, M. A. et al. Somatic mutation in single human neurons tracks developmental and transcriptional history. *Science* **350**, 94–98 (2015).

[ADS](#) [CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

13. 13.

Lodato, M. A. et al. Aging and neurodegeneration are associated with increased mutations in single human neurons. *Science* **359**, 555–559 (2018).

[ADS](#) [CAS](#) [PubMed](#) [Google Scholar](#)

14. 14.

Brazhnik, K. et al. Single-cell analysis reveals different age-related somatic mutation profiles between stem and differentiated cells in human liver. *Sci. Adv.* **6**, eaax2659 (2020).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

15. 15.

Xing, D., Tan, L., Chang, C. H., Li, H. & Xie, X. S. Accurate SNV detection in single cells by transposon-based whole-genome amplification of complementary strands. *Proc. Natl Acad. Sci. USA* **118**, e2013106118 (2021).

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

16. 16.

Petljak, M. et al. Characterizing mutational signatures in human cancer cell lines reveals episodic APOBEC mutagenesis. *Cell* **176**, 1282–1294.e20 (2019).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

17. 17.

Salk, J. J., Schmitt, M. W. & Loeb, L. A. Enhancing the accuracy of next-generation sequencing for detecting rare and subclonal mutations. *Nat. Rev. Genet.* **19**, 269–285 (2018).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

18. 18.

Schmitt, M. W. et al. Detection of ultra-rare mutations by next-generation sequencing. *Proc. Natl Acad. Sci. USA* **109**, 14508–14513 (2012).

[ADS](#) [CAS](#) [PubMed](#) [Google Scholar](#)

19. 19.

Kennedy, S. R. et al. Detecting ultralow-frequency mutations by duplex sequencing. *Nat. Protocols* **9**, 2586–2606 (2014).

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

20. 20.

Hoang, M. L. et al. Genome-wide quantification of rare somatic mutations in normal human tissues using massively parallel sequencing. *Proc. Natl Acad. Sci. USA* **113**, 9846–9851 (2016).

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

21. 21.

You, X. et al. Detection of genome-wide low-frequency mutations with paired-end and complementary consensus sequencing (PECC-seq) revealed end-repair-derived artifacts as residual errors. *Arch. Toxicol.* **94**, 3475–3485 (2020).

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

22. 22.

Costello, M. et al. Discovery and characterization of artifactual mutations in deep coverage targeted capture sequencing data due to oxidative DNA damage during sample preparation. *Nucleic Acids Res.* **41**, e67 (2013).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

23. 23.

Kong, A. et al. Rate of de novo mutations and the importance of father's age to disease risk. *Nature* **488**, 471–475 (2012).

[ADS](#) [CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

24. 24.

Rahbari, R. et al. Timing, rates and spectra of human germline mutation. *Nat. Genet.* **48**, 126–133 (2016).

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

25. 25.

Wyles, S. P., Brandt, E. B. & Nelson, T. J. Stem cells: the pursuit of genomic stability. *Int. J. Mol. Sci.* **15**, 20948–20967 (2014).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

26. 26.

Lee-Six, H. et al. Population dynamics of normal human blood inferred from somatic mutations. *Nature* **561**, 473–478 (2018).

[ADS](#) [CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

27. 27.

Nicholson, A. M. et al. Fixation and spread of somatic mutations in adult human colonic epithelium. *Cell Stem Cell* **22**, 909–918.e8 (2018).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

28. 28.

Pleguezuelos-Manzano, C. et al. Mutational signature in colorectal cancer caused by genotoxic pks<sup>+</sup> *E. coli*. *Nature* **580**, 269–273 (2020).

[ADS](#) [CAS](#) [PubMed](#) [Google Scholar](#)

29. 29.

Poduri, A., Evrony, G. D., Cai, X. & Walsh, C. A. Somatic mutation, genomic variation, and neurological disease. *Science* **341**, 1237758 (2013).

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

30. 30.

Alexandrov, L. B. et al. The repertoire of mutational signatures in human cancer. *Nature* **578**, 94–101 (2020).

[ADS](#) [CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

31. 31.

Rheinbay, E. et al. Analyses of non-coding somatic drivers in 2,658 cancer whole genomes. *Nature* **578**, 102–111 (2020).

[ADS](#) [CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

32. 32.

Gabella, G. Cells of visceral smooth muscles. *J. Smooth Muscle Res.* **48**, 65–95 (2012).

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

33. 33.

Yoshida, K. et al. Tobacco smoking and somatic mutations in human bronchial epithelium. *Nature* **578**, 266–272 (2020).

[ADS](#) [CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

34. 34.

Lawson, A. R. J. et al. Extensive heterogeneity in somatic mutation and selection in the human bladder. *Science* **370**, 75–82 (2020).

[ADS](#) [CAS](#) [PubMed](#) [Google Scholar](#)

35. 35.

Gao, Z., Wyman, M. J., Sella, G. & Przeworski, M. Interpreting the dependence of mutation rates on age and time. *PLoS Biol.* **14**, e1002355 (2016).

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

36. 36.

Kucab, J. E. et al. A compendium of mutational signatures of environmental agents. *Cell* **177**, 821–836 (2019).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

37. 37.

Matsumura, S. et al. Genome-wide somatic mutation analysis via Hawk-seq™ reveals mutation profiles associated with chemical mutagens. *Arch. Toxicol.* **93**, 2689–2701 (2019).

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

38. 38.

Ellis, P. et al. Reliable detection of somatic mutations in solid tissues by laser-capture microdissection and low-input DNA sequencing. *Nat. Protocols* **16**, 841–871 (2021).

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

39. 39.

Olafsson, S. et al. Somatic evolution in non-neoplastic IBD-affected colon. *Cell* **182**, 672–684 (2020).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

40. 40.

Krishnaswami, S. R. et al. Using single nuclei for RNA-seq to capture the transcriptome of postmortem neurons. *Nat. Protocols* **11**, 499–524 (2016).

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

41. 41.

Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. Preprint at <https://arxiv.org/abs/1303.3997> (2013).

42. 42.

Tischler, G. & Leonard, S. biobambam: tools for read pair collation based algorithms on BAM files. *Source Code Biol. Med.* **9**, 13 (2014).

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

43. 43.

Gerstung, M. et al. Reliable detection of subclonal single-nucleotide variants in tumour cell populations. *Nat. Commun.* **3**, 811 (2012).

[ADS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

44. 44.

Karczewski, K. J. et al. The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature* **581**, 434–443 (2020).

[ADS](#) [CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

45. 45.

The 1000 Genomes Project Consortium. A global reference for human genetic variation. *Nature* **526**, 68–74 (2015).

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

46. 46.

Zhang, F. et al. Ancestry-agnostic estimation of DNA sample contamination from sequence reads. *Genome Res.* **30**, 185–194 (2020).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

47. 47.

Benjamini, Y. & Speed, T. P. Summarizing and correcting the GC content bias in high-throughput sequencing. *Nucleic Acids Res.* **40**, e72 (2012).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

48. 48.

Robinson, P. S. et al. Elevated somatic mutation burdens in normal human cells due to defective DNA polymerases. Preprint at <https://doi.org/10.1101/2020.06.23.167668> (2020).

49. 49.

Jones, D. et al. cgpCaVEManWrapper: simple execution of CaVEMan in order to detect somatic single nucleotide variants in NGS data. *Curr. Protoc. Bioinformatics* **56**, 15.10.1–15.10.18 (2016).

[Google Scholar](#)

50. 50.

Raine, K. M. et al. Cgppindel: identifying somatically acquired insertion and deletion events from paired end sequencing. *Curr Protoc Bioinformatics* **52**, 15.17.1–15.17.12 (2015).

[Google Scholar](#)

51. 51.

Hoang, D. T. et al. MPBoot: fast phylogenetic maximum parsimony tree inference and bootstrap approximation. *BMC Evol. Biol.* **18**, 11 (2018).

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

52. 52.

Gori, K. & Baez-Ortega, A. sigfit: flexible Bayesian inference of mutational signatures. Preprint at <https://doi.org/10.1101/372896> (2020).

53. 53.

Lensing S. V. et al. Somatic mutation landscapes at single-molecule resolution. *Protocol Exchange* <https://doi.org/10.21203/rs.3.pex-1298/v1> (2021).

54. 54.

Lawrence, M. et al. Software for computing and annotating genomic ranges. *PLOS Comput. Biol.* **9**, e1003118 (2013).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

55. 55.

Gerstung, M., Papaemmanuil, E. & Campbell, P. J. Subclonal variant calling with multiple samples and prior knowledge. *Bioinformatics* **30**, 1198–1204 (2014).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

56. 56.

Li, H. et al. The sequence alignment/map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).

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

57. 57.

Li, H. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* **27**, 2987–2993 (2011).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

58. 58.

Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

59. 59.

Kundaje, A. et al. Integrative analysis of 111 reference human epigenomes. *Nature* **518**, 317–330 (2015).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

[Download references](#)

## Acknowledgements

We are grateful to the live donors and the families of the deceased organ transplant donors. We thank L. Anderson, K. Roberts, C. Latimer, Q. Lin, members of the CGP-lab, R. Vicario, F. Geissmann, N. Angelopoulos, G. Tischler, T. Bellerby, M. Abascal and K. Chatterjee for assistance in the development of NanoSeq or with this manuscript; all NIHR BioResource Centre Cambridge volunteers for participation; the NIHR BioResource Centre Cambridge and staff for their contribution; the National Institute for Health Research and NHS Blood and Transplant; the Cambridge Blood and Stem Cell Biobank for sample donation and support of this work; the Cambridge Brain Bank for sample donation; and the participants and local coordinators at the TwinsUK study. This research was supported by the Cambridge NIHR BRC Cell Phenotyping Hub. I.M. is funded by Cancer Research UK (C57387/A21777) and the Wellcome Trust. P.J.C. is a Wellcome Trust Senior Clinical Fellow. R.R. is a recipient of a CRUK Career Development fellowship (C66259/A27114). E.L. is supported by a Wellcome/Royal Society Sir Henry Dale Fellowship (grant number 107630/Z/15/Z), the European Hematology Association, BBSRC and by core funding from Wellcome (grant number 203151/Z/16/Z) and MRC to the Wellcome-MRC Cambridge Stem Cell Institute. D.G.K. is supported by

a Bloodwise Bennett Fellowship (15008), the Bill and Melinda Gates Foundation (INV-002189) and an ERC Starting Grant (ERC-2016-STG-715371). The TwinsUK study was funded by the Wellcome Trust and European Community's Seventh Framework Programme (FP7/2007-2013). The TwinsUK study also receives support from the National Institute for Health Research (NIHR)-funded BioResource, Clinical Research Facility and Biomedical Research Centre based at Guy's and St Thomas' NHS Foundation Trust in partnership with King's College London. The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health & Social Care.

## Author information

### Author notes

1. Peter Ellis

Present address: Inivata, Babraham Research Campus, Cambridge, UK

2. Robert J. Osborne

Present address: Biofidelity, Cambridge Science Park, Cambridge, UK

3. These authors contributed equally: Luke M. R. Harvey, Emily Mitchell, Andrew R. J. Lawson, Stefanie V. Lensing, Peter Ellis

## Affiliations

1. Wellcome Sanger Institute, Hinxton, UK

Federico Abascal, Luke M. R. Harvey, Emily Mitchell, Andrew R. J. Lawson, Stefanie V. Lensing, Peter Ellis, Andrew J. C. Russell, Raul E. Alcantara, Adrian Baez-Ortega, Yichen Wang, Eugene Jing Kwa, Henry Lee-Six, Alex Cagan, Tim H. H. Coorens, Michael Spencer Chapman, Sigurgeir Olafsson, Steven Leonard, David Jones, Heather E. Machado, Michael R. Stratton, Raheleh Rahbari, Peter J. Campbell, Robert J. Osborne & Iñigo Martincorena

2. Wellcome–MRC Cambridge Stem Cell Institute, Cambridge Biomedical Campus, Cambridge, UK

Emily Mitchell, Megan Davies, Nina F. Øbro, David G. Kent & Elisa Laurenti

3. Department of Haematology, University of Cambridge, Cambridge, UK

Nina F. Øbro, Krishnaa T. Mahubani, Elisa Laurenti & Peter J. Campbell

4. Department of Surgery, University of Cambridge, Cambridge, UK

Krishnaa T. Mahubani & Kourosh Saeb-Parsy

5. NIHR Cambridge Biomedical Research Centre, Cambridge Biomedical Campus, Cambridge, UK

Krishnaa T. Mahubani & Kourosh Saeb-Parsy

6. Cambridge Brain Bank, Division of the Human Research Tissue Bank, Addenbrooke's Hospital, Cambridge, UK

Kieren Allinson

7. European Molecular Biology Laboratory, European Bioinformatics Institute (EMBL-EBI), Hinxton, UK

Moritz Gerstung

8. York Biomedical Research Institute, Department of Biology, University of York, York, UK

David G. Kent

## Authors

1. Federico Abascal

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

2. Luke M. R. Harvey

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

3. Emily Mitchell

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

4. Andrew R. J. Lawson

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

5. Stefanie V. Lensing

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

6. Peter Ellis

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

7. Andrew J. C. Russell

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

8. Raul E. Alcantara

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

9. Adrian Baez-Ortega

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

10. Yichen Wang

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

11. Eugene Jing Kwa

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

12. Henry Lee-Six

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

13. Alex Cagan

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

14. Tim H. H. Coorens

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

15. Michael Spencer Chapman

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

16. Sigurgeir Olafsson

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

17. Steven Leonard

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

18. David Jones

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

19. Heather E. Machado

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

20. Megan Davies

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

21. Nina F. Øbro

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

22. Krishnaa T. Mahubani

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

23. Kieren Allinson

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

24. Moritz Gerstung

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

25. Kourosh Saeb-Parsy

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

26. David G. Kent

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

27. Elisa Laurenti

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

28. Michael R. Stratton

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

29. Raheleh Rahbari

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

30. Peter J. Campbell

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

31. Robert J. Osborne

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

32. Iñigo Martincorena

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

## Contributions

R.J.O., F.A. and I.M. conceived the project. I.M., P.J.C., R.R. and M.R.S. supervised the project. F.A., R.J.O., E.M. and I.M. wrote the manuscript; all authors reviewed and edited the manuscript. R.J.O. led the development of the protocol with help from F.A., A.R.J.L., P.E., S.V.L. and I.M. R.J.O. and F.A. developed the bioinformatics pipeline with help from R.E.A., S.L. and D.J. F.A. led the analysis of the data with help from A.R.J.L., I.M., A.B.-O., Y.W., L.M.R.H., E.J.K., T.H.H.C., M.S.C. and M.G. E.M. performed the HSC/MPP experiments. L.M.R.H. and A.J.C.R. performed the cell sorting of neuronal nuclei. A.R.J.L. and A.C. performed laser microdissection. E.M., N.F.Ø., H.E.M., M.D., D.G.K., E.L., K.T.M., K.S.-P., K.A., R.R., H.L.-S. and S.O. collected and processed samples. E.M., E.L., M.G. and D.G.K. assisted in the interpretation of blood data.

## Corresponding authors

Correspondence to [Robert J. Osborne](#) or [Iñigo Martincorena](#).

## Ethics declarations

## Competing interests

A patent application on NanoSeq has been filed that includes R.J.O., F.A. and I.M.

## Additional information

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

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

## Extended data figures and tables

### Extended Data Fig. 1 Substitution imbalances and impact of A-tailing.

**a, b**, Imbalances in the distribution of the six complementary substitutions (for example, G>T versus C>A) across read positions in BotSeqS (**a**) and NanoSeq (**b**). **c**, Origin of G>T over C>A mutation call imbalances in standard sequencing<sup>22</sup>. **d**, Origin of imbalances in duplex sequencing/BotSeqS as a result of end repair during library preparation. **e**, Single-strand consensus calls for pyrimidine (top) and purine (bottom) substitutions for the standard BotSeqS (left) protocol and for NanoSeq with standard (middle) and modified (right) A-tailing protocols. For example, C>T changes are shown at the top, whereas the complementary G>A changes are shown at the bottom. By using ddBTPs, C>A, G>A and T>A errors are reduced, lowering the risk of false-positive double-strand consensus calls.

### Extended Data Fig. 2 BotSeqS errors as a function of read-end trimming.

**a**, BotSeqS estimated burden for the granulocyte sample shown in Fig. 2c applying different extents of trimming to the 5' ends of reads. Even with extensive trimming we predict at least 600 artifactual mutation calls per diploid genome. **b**, Substitution imbalances are observed deep into the reads and cannot be avoided with read trimming. Imbalances vary from experiment to experiment, as a consequence of DNA damage in the DNA source or during library preparation (Supplementary Note 1). **c**, Substitution profiles including the reference profile from single-cell-derived blood colonies and three BotSeqS profiles after trimming of 20, 40 and 60 bp from the 5' end of reads (in addition to 15 bp trimming of the 3' end). The text in the figure indicates the observed and expected cosine similarities (Methods) to the reference profile. C>A and C>G errors in BotSeqS remain after extensive trimming.

### Extended Data Fig. 3 Mung Bean NanoSeq.

**a**, Estimated number of mutations per cord blood cell. Poisson 95% confidence intervals are shown as lines. The red dotted line shows the number of mutations per cord blood cell estimated with the restriction enzyme NanoSeq protocol, with Poisson 95% confidence intervals shown as a red shade. In contrast to Fig. 1f, we did not apply the correction for missing embryonic mutations because here we are comparing two protocols that are equally affected by this limitation. **b**, Substitution profiles for the standard end-repair protocol (BotSeqS) and for Mung Bean, showing the cosine similarities with the reference profile (Fig. 1c).

#### Extended Data Fig. 4 Optimization of duplicate rates, DNA input requirements and estimation of human contamination.

**a**, Relationship between sequencing yield, library complexity, duplicate rates and efficiency, based on a truncated Poisson model (Methods). Left, duplicate rate as a function of the sequencing ratio (sequencing reads/DNA fragments in the library). Middle, efficiency (measured as bases called with duplex coverage/bases sequenced) as a function of the duplicate rate. Right, efficiency as a function of sequencing ratio. **b**, Library yield as fmol per 25 µl as a function of the amount of input DNA in ng. **c**, Empirical relationship between the estimated fmol in library (measured by qPCR) and the number of unique molecules in the library estimated with Picard tools (Lander–Waterman equation) for our choice of restriction enzyme and fragment size selection (250–500 bp). **d**, Empirical relationship between duplicate rates and efficiency of the method, measured as duplex bases called/number of bases sequenced (that is, the number of paired-end reads multiplied by 300). The maximum efficiency (around 0.04) is lower than the maximum analytical expectation (0.12; middle panel in **a**) because of the trimming of read ends (barcodes, restriction sites and 8 bp from each end) and the strict filters that we apply to consider a site callable. **e**, VerifyBamId contamination estimates for different amounts of simulated contamination from individuals of different ancestry. **f**, Contamination simulation using two NanoSeq samples to contaminate each other.

#### Extended Data Fig. 5 Correction of standard (CaVEMan-based) mutation burden estimates and validation of NanoSeq

## indel calling.

**a**, Comparison of the mutation burden estimates in regions of the genome with at least 20 $\times$  coverage (*c*) to the trinucleotide-context-corrected mutation burdens in the subset of *c* covered by NanoSeq and passing all NanoSeq filters. **b**, Ratio between the rates shown in **a**, showing that the corrected burden is approximately 20% higher than the uncorrected burden; box plots show the interquartile range, median and 95% confidence interval for the median. **c**, Comparison of indel rates between cord blood colonies (indels were called with the Pindel algorithm) and granulocytes from neonates (NanoSeq pipeline), showing Poisson 95% confidence intervals. Given the sparsity of indel calls in cord blood, data from different colonies ( $n = 100$ ) and granulocytes ( $n = 2$  donors, one of them with 5 replicates) were combined into single point estimates. **d**, The top two panels show the high similarity between the NanoSeq and Pindel indel profiles for a bladder tumour; the bottom two profiles show the indel spectra in blood from *POLE* and *POLD1* germline mutation carriers, which are very similar to previously reported profiles<sup>48</sup>.

## Extended Data Fig. 6 Cell sorting of HSC/MPPs and colon histology.

**a**, Gating strategy for the isolation of HSC/MPPs from a representative bone marrow sample. Text above the plots indicates the population depicted. Text inside the plots indicates the name of the gates shown in pink. The CD34 $^+$ CD38 $^-$  population is defined as the bottom 20% CD38 $^-$  as shown. For all initial samples (bone marrow, peripheral blood and cord blood), the index sorted population is the ‘HSC pool’ gate. Cell population abundance differed between samples but typically viable cells were 60–90% of total cells and singlets were 98–99% of viable cells. Live cells were 90–99% of viable cells and myeloid cells were 15–50% of live cells. CD34 $^+$  cells were typically 1–15% of myeloid cells. **b, c**, Colon histology sections showing microbiopsied areas of colonic epithelium and smooth muscle for donors PD34200 (**b**) and PD37449 (**c**). For donor PD34200, a single crypt, a pool of six crypts and two smooth muscle areas were sequenced. For donor PD37449, the two single crypts and the pool of six crypts were

sequenced. The burden estimates for these microbiopsies are shown in Figs. [2d](#), [3j](#), [k](#). **d**, Substitution profiles for colonic crypts from the three donors in Fig. [2d](#) and cosine similarities to profiles obtained with standard methods.

### **Extended Data Fig. 7 Neuron nuclei sorting, comparison to single-cell data and accumulation of mutations with age.**

**a**, Gating strategy for the isolation of neuronal nuclei from frontal cortex. Nuclei were sorted by FACS using an Influx cell sorter (BD Biosciences) with a 100- $\mu\text{m}$  nozzle. For each sample an unstained control was used to help to determine the NeuN<sup>+</sup> population. The text above each column indicates the population depicted and the text inside the plots indicates the population of the gates highlighted in black. Sorting results varied among samples, with 1–60% passing the DAPI gate and, of these, 2–53% passing a conservative NeuN<sup>+</sup> gate. **b**, Substitution profiles for all mutations detected in neurons with SNP-phased error-corrected single-cell sequencing data from a previously published study<sup>[13](#)</sup> (top) and with NanoSeq (middle). Bottom, a signature specific to single-cell sequencing data is shown (scF signature from a previous publication<sup>[16](#)</sup>). **c**, Mutational signatures extracted from a previously published study<sup>[13](#)</sup>, showing their relative contributions in the published dataset. These signatures were obtained using sigfit (Methods) on publicly available mutation calls and are referred to as LDA, LDB and LDC. Note the high similarity between the NanoSeq full spectrum for neurons and LDA (cosine similarity 0.96), and between scF and LDB (cosine similarity 0.97). **d**, Predicted contribution of LDA, LDB and LDC to each of the previously sequenced neurons<sup>[13](#)</sup>. **e**, Accumulation of mutations attributed to NanoSeq signatures A, B and C with age in healthy donors and in individuals with Alzheimer’s disease. **f**, Accumulation of mutations attributed to NanoSeq signatures A, B and C in smooth muscle from bladder and colon.

### **Extended Data Fig. 8 Normalized substitution spectra across different genomic regions.**

**a**, Substitution spectra for neurons, granulocytes, smooth muscle and colonic crypts in chromatin states associated to transcription (states E4 and

E5 in ENCODE) and inactive DNA (E9 and E15). Chromatin states were obtained from ENCODE<sup>59</sup>, using the following epigenomes: E073 (frontal cortex), E030 (granulocytes), E076 (smooth muscle) and E075 (colonic mucosa). To enable the direct comparison of spectra across genomic regions with different trinucleotide frequencies, the profiles have been normalized to the genomic trinucleotide frequencies (Methods). **b**, Transcriptional strand asymmetries in neurons, granulocytes, smooth muscle and colonic crypts. **c**, Transcriptional strand asymmetries in neurons in quartiles of gene expression.

### Extended Data Fig. 9 Additional substitution and indel spectra.

**a**, NanoSeq mutational spectrum for neurons corrected for trinucleotide frequency in the callable genome. Unlike the usual representation, which shows unnormalized rates, this representation shows mutation rates per available trinucleotide. **b**, Previously published LDA signature<sup>13</sup> normalized to trinucleotide frequency in the genome also reveals high C>T rates at CpG dinucleotides. This observation from single-cell data suggests that the high C>T rates at CpG sites in NanoSeq neuron data (**a**) are not caused by contamination of NeuN<sup>+</sup> pools with glia or other cells. **c**, Indel profiles of granulocytes (top) and colonic crypts without the colibactin signature (bottom). **d**, Indel profiles for the 250 most highly expressed genes in the PCAWG liver hepatocellular carcinoma data<sup>31</sup>.

### Extended Data Fig. 10 Smooth muscle.

**a**, Histology of bladder smooth muscle showing two sections from donor PD40842; only one of the two sections was sequenced using NanoSeq. **b**, Number of mutations detected with CaVEMan in different smooth muscle sections processed with our standard microdissection sequencing protocol<sup>38</sup>. The orange dots show the expected mutation burdens (with 95% confidence intervals) for these sections based on the donor age and the regression model shown in Fig. 3j. **c**, Distribution of VAFs for each of the smooth muscle sections using standard whole-genome sequencing. Box plots show the interquartile range, median, 95% confidence interval for the median (notches), and outliers (black dots).

# Supplementary information

## Supplementary Information

This file contains Supplementary Notes 1-10, including Supplementary Figures 1-5 and Supplementary References.

## Reporting Summary

## Supplementary Tables

This file contains Supplementary Tables 1-8. Supplementary Table 1 lists samples used in this study and corresponding data availability. Supplementary Table 2 displays sequencing yields for NanoSeq/BotSeqS DNA libraries. Supplementary Table 3 shows *in silico* restriction enzyme digestion of the human genome. Supplementary Table 4 displays substitution and indel rates. Supplementary Table 5 shows substitution calls (NanoSeq protocol). Supplementary Table 6 shows indel calls (NanoSeq protocol). Supplementary Table 7 displays trinucleotide substitution profiles and Supplementary Table 8 shows Linear regression models.

## Peer Review File

## Rights and permissions

### Reprints and Permissions

## About this article



Check for  
updates

## Cite this article

Abascal, F., Harvey, L.M.R., Mitchell, E. *et al.* Somatic mutation landscapes at single-molecule resolution. *Nature* **593**, 405–410 (2021). <https://doi.org/10.1038/s41586-021-03477-4>

## [Download citation](#)

- Received: 13 November 2020
- Accepted: 22 March 2021
- Published: 28 April 2021
- Issue Date: 20 May 2021
- DOI: <https://doi.org/10.1038/s41586-021-03477-4>

## Comments

By submitting a comment you agree to abide by our [Terms](#) and [Community Guidelines](#). If you find something abusive or that does not comply with our terms or guidelines please flag it as inappropriate.

[Access through your institution](#)

[Change institution](#)

[Buy or subscribe](#)

Advertisement

---

This article was downloaded by **calibre** from <https://www.nature.com/articles/s41586-021-03477-4>

Mouse prefrontal cortex represents learned rules for categorization

[Download PDF](#)

- Article
- Open Access
- [Published: 21 April 2021](#)

# Mouse prefrontal cortex represents learned rules for categorization

- [Sandra Reinert](#)<sup>1,2</sup>,
- [Mark Hübener](#) [ORCID: orcid.org/0000-0001-8367-9132](#)<sup>1</sup>,
- [Tobias Bonhoeffer](#) [ORCID: orcid.org/0000-0001-7897-6634](#)<sup>1</sup> &
- [Pieter M. Goltstein](#) [ORCID: orcid.org/0000-0001-8895-0841](#)<sup>1</sup>

[Nature](#) volume 593, pages 411–417 (2021) [Cite this article](#)

- 12k Accesses
- 251 Altmetric
- [Metrics details](#)

## Subjects

- [Cortex](#)
- [Learning and memory](#)
- [Long-term memory](#)
- [Neural circuits](#)
- [Operant learning](#)

## Abstract

The ability to categorize sensory stimuli is crucial for an animal's survival in a complex environment. Memorizing categories instead of individual exemplars enables greater behavioural flexibility and is computationally advantageous. Neurons that show category selectivity have been found in several areas of the mammalian neocortex<sup>1,2,3,4</sup>, but the prefrontal cortex seems to have a prominent role<sup>4,5</sup> in this

context. Specifically, in primates that are extensively trained on a categorization task, neurons in the prefrontal cortex rapidly and flexibly represent learned categories<sup>6,7</sup>. However, how these representations first emerge in naive animals remains unexplored, leaving it unclear whether flexible representations are gradually built up as part of semantic memory or assigned more or less instantly during task execution<sup>8,9</sup>. Here we investigate the formation of a neuronal category representation throughout the entire learning process by repeatedly imaging individual cells in the mouse medial prefrontal cortex. We show that mice readily learn rule-based categorization and generalize to novel stimuli. Over the course of learning, neurons in the prefrontal cortex display distinct dynamics in acquiring category selectivity and are differentially engaged during a later switch in rules. A subset of neurons selectively and uniquely respond to categories and reflect generalization behaviour. Thus, a category representation in the mouse prefrontal cortex is gradually acquired during learning rather than recruited ad hoc. This gradual process suggests that neurons in the medial prefrontal cortex are part of a specific semantic memory for visual categories.

[Download PDF](#)

## Main

We trained head-fixed mice ( $n = 11$ ) in a ‘Go’/‘NoGo’ rule-based categorization task (Fig. 1a,b) to sort visual stimuli into two groups. Stimuli were 36 sinusoidal gratings, each with a specific combination of two stimulus features: orientation and spatial frequency. At any time, one rule determined the relevant feature for categorization (that is, the active rule; for example, assigning category identity of a stimulus based on orientation)<sup>10,11</sup> (Extended Data Fig. 1). First, mice learned to discriminate two exemplar stimuli according to the active rule. All mice achieved more than 66% correct Go choices within 10 to 40 sessions, showing considerable variability in the rate of learning. We then introduced categories by stepwise addition of stimuli to the task, up to a set of 18 different gratings that varied along both feature dimensions, orientation and spatial frequency (Extended Data Fig. 1b,c). Mice integrated the newly introduced stimuli within 1 to 2 sessions and they maintained a sensitivity index  $d$  of  $>1$  (Fig. 1c,d, Extended Data Figs. 1d, 2).

**Fig. 1: Mice learn rules to categorize visual stimuli and generalize to novel stimuli.**

---

 **figure1**

**a**, Schematic of behavioural training setup. **b**, Schematic of trial structure in the Go/NoGo task. ITI, inter-trial interval; Stim./resp., stimulus presentation/response window. **c**, Performance ( $d'$ ) of 11 mice in each training session. Individual traces aligned to criterion (66% of correct trials). The dashed line indicates chance level ( $d' = 0$ ). Crosses denote sessions with two-photon imaging (T1–T8). The spread in performance after T2 is due to day-to-day variability rather than mouse-to-mouse variability. TP, time point. **d**, Fraction of Go choices per stimulus of an example mouse at each time point (of two-photon imaging) until the presentation of all 36 stimuli of rule 1 (generalization; T5). **e**, Performance ( $d'$ ) for rule 1 (T5), for experienced (Exp.) compared to novel (Nov.) stimuli.  $P = 0.50$ , two-tailed paired-samples  $t$ -test ( $n = 11$  mice). Grey lines denote individual mice. Data are mean  $\pm$  s.e.m. **f**, Number of training sessions until criterion (66% correct, exemplar stimuli). Bars indicate mean across mice, dots are individual mice (green denotes the orientation rule; orange denotes spatial frequency rule). Rule 2 is learned significantly faster than rule 1.  
 $P = 9.77 \times 10^{-4}$ , two-tailed Wilcoxon matched-pairs signed-rank (WMPSR) test ( $n = 11$  mice). **g**, As in **d**, for rule 2 of the same mouse. **h**, As in **e** for rule 2 (T8).  $d'$  did not differ significantly between novel stimuli and stimuli experienced with rule 2.  $P = 0.09$ , two-tailed paired-samples  $t$ -test ( $n = 10$  mice). **i**, Schematics specifying the distance of stimuli to the boundary. **j**, Psychometric curves showing the fraction of Go

choices along the relevant (black) and irrelevant (blue) dimension of rule 1 at T1, T5 and T8. Left:  $P_{\text{relevant T1}} = 0.36$ ,  $P_{\text{irrelevant T1}} = 0.77$ ; middle: \*\*\* $P_{\text{relevant T5}} = 1.73 \times 10^{-6}$ ,  $P_{\text{irrelevant T5}} = 0.09$ ; right:  $P_{\text{relevant T5}} = 0.73$ , \*\*\* $P_{\text{irrelevant T5}}$  (relevant T8) =  $1.73 \times 10^{-6}$ , two-tailed WMPSR test, Bonferroni-corrected for two comparisons ( $n = 10$  mice). Categorization performance was not affected by the order in which mice were trained on orientation and spatial frequency rules. Data are mean  $\pm$  s.e.m. across mice; for individual mice, see Extended Data Fig. 2. NS, not significant.

[Source data](#)

[Full size image](#)

## Mice learn to categorize visual stimuli

To determine whether mice had indeed learned categorization, we tested a characteristic feature of category learning, rapid generalization to novel stimuli<sup>10,11,12,13</sup>. Mice were presented with 18 novel grating stimuli in addition to the 18 experienced ones. All mice were able to generalize the learned rule to novel stimuli upon their first presentation (time point 5, T5) (Fig. 1d, Extended Data Fig. 3a), performing equally well on novel and experienced stimuli (Fig. 1e, Extended Data Fig. 3b).

Because rule-switching is key to rule-based categorization<sup>11,14,15</sup>, our stimulus set was designed to allow for testing this aspect. Thus, after learning the first rule, mice were required to group the same stimuli into new categories according to a new rule, by making the previously irrelevant stimulus feature (for example, spatial frequency) relevant and the relevant one (for example, orientation) irrelevant. All mice learned to discriminate two exemplar stimuli for rule 2 considerably faster than during initial learning (Fig. 1f, Extended Data Fig. 1e-h). After the mice had learned to categorize a set of 18 stimuli according to rule 2, they were able to generalize to the 18 stimuli they had so far experienced only with rule 1 (Fig. 1g,h, Extended Data Fig. 3c-f). We tested whether there were any residual effects of the former rule on the categorization behaviour of the mice by comparing the influence of each stimulus feature (Fig. 1i) on the choices of the mice before learning (time point T1) and after learning each rule (T5 and T8). Untrained mice showed no effect of either stimulus feature on categorization (Fig. 1j, left). Trained mice only based categorization on the stimulus feature relevant to the active rule; the irrelevant feature showed no effect (Fig. 1j, middle, right; for a detailed analysis of categorization near the boundary see Extended Data Fig. 3g-l). In summary, all mice learned discriminating categories on the basis of two different rules, and they generalized these rules to novel stimuli, probably by selectively attending<sup>16</sup>

to the relevant stimulus feature. Having established this training paradigm, we began tracking neuronal correlates of rule-based categories throughout learning.

## mPFC contains category-selective neurons

The prefrontal cortex (PFC) in primates and rodents is important for cognitive functions such as categorization<sup>6,7,16,17</sup>, rule learning<sup>1,18,19</sup> and cognitive flexibility<sup>20,21,22</sup>, even though the functional and anatomical analogy of this region across species is still debated<sup>23,24,25,26,27,28</sup>. Earlier studies in primates have described individual PFC neurons coding for visual categories<sup>3,6,7</sup>. Encouraged by this finding, we tested whether the mouse medial PFC (mPFC) contained neurons that reflected the ability of the mouse to categorize visual stimuli as described above. To this end, we chronically monitored neuronal activity in cortical layer 2/3 using two-photon calcium imaging through a microprism implant inserted between the two hemispheres, which enabled optical access to mPFC<sup>29</sup> (Fig. 2a–c, Extended Data Fig. 4). We measured neuronal activity of individual cells while the mice performed the task ( $d'$  ranging from 0.7 to 3.6; for imaging time points and precise trial structure, see Fig. 1b,c, Extended Data Fig. 1a). In naive mice (time point T1), mPFC neurons did not respond to visual stimuli (Fig. 2b,d, Extended Data Fig. 5), but some of these initially non-selective cells clearly showed category selectivity after learning (T5, rule 1) (Fig. 2c,e, Extended Data Fig. 5; neural correlates of other task-related aspects are described below).

**Fig. 2: Single neurons in the mouse mPFC develop category-selective responses.**

---

 **figure2**

**a**, Schematic depicting virus injection into the mPFC and two-photon imaging through a prism implant between the hemispheres, adapted from ref. [29](#). AAV, adeno-associated virus. **b**, Example field of view before learning (T1). Left, pseudo-coloured GCaMP6m (green) and mRuby2 (red) fluorescence. Right, hue, saturation and lightness (HLS) map. Hue: preferred category; lightness: response amplitude; saturation: category selectivity. Scale bar, 30  $\mu$ m. **c**, As in **b**, after learning (T5). **d**, Left, stimulus space. Middle, average response to stimuli of the cell highlighted in **b** ( $\Delta F/F$ , green: Go category, red: NoGo category). L, average time of first lick. R, average time of reward delivery. Right, inferred spike rate. AU, arbitrary units. **e**, As in **d**, after learning, showing selective responsiveness to Go category stimuli. **f**, Left, CTI of all cells in the example field of view before and after learning. Red, example cell in **b**. Grey line denotes the threshold used for further analyses (CTI > 0.1). Right, cross-validated performance of a Bayesian decoder predicting the category of the presented stimulus. CTI correlates with decoding performance.  $P = 2.2 \times 10^{-10}$ , rho = 0.41, Spearman's correlation ( $n = 213$  category-selective neurons, CTI > 0.1). **g**, Percentage of category-selective cells for rule 1. T1, before learning, rule 1 active. T5, after learning, rule 1. T8, after learning, rule 2.  $P_{T1-T5} = 0.006$ ,  $P_{T1-T8} = 0.25$ ,  $P_{T5-T8} = 0.004$ , two-tailed WMPSR test, Bonferroni-corrected for three comparisons ( $n_{\text{mice}} = 10$ ,  $n_{\text{neurons}} = 2,306$ ). **h**, As in **g**, for rule 2.  $P_{T1-T5} = 0.75$ ,  $P_{T1-T8} = 0.004$ ,  $P_{T5-T8} = 0.004$ , two-tailed WMPSR test, Bonferroni-corrected for three comparisons ( $n = 10$  mice). **i**, Bayesian decoding performance as in **f**, for all mice. Data are shown separately for populations of low (red) and high (black) CTI cells. Light grey denotes

individual mice; dark grey denotes average performance after shuffling stimulus categories.  $P_{1 \text{ neuron}} = 0.005$ ,  $P_{2-8 \text{ neurons}} = 0.01$ , two-tailed WMPSR test, high versus low CTI cells, Bonferroni-corrected for two comparisons ( $n_{1 \text{ neuron}} = 10$  mice,  $n_{2-8 \text{ neurons}} = 8$  mice). Data are mean  $\pm$  s.e.m. across mice; for individual mice see Extended Data Figs. [5](#), [6](#).

[Source data](#)

[Full size image](#)

We quantified the category selectivity of individual cells using the category-tuning index (CTI)<sup>30</sup>, with values close to 1 indicating strong differences in activity between, but not within, categories, and a value of 0 indicating no difference in the firing rate between and within categories. We defined neurons with a CTI value above 0.1 as category-selective (Methods), and verified that these cells reliably encoded categories using cross-validated Bayesian decoding (Fig. [2f,i](#)). In naive mice, hardly any cells exceeded this threshold, whereas after learning, a substantial fraction of neurons in the mPFC showed category selectivity (before:  $0.03\% \pm 0.03\%$ , after:  $9.8\% \pm 2.2\%$  (mean  $\pm$  s.e.m.)) (Fig. [2g](#), Extended Data Fig. [6a](#)).

After having learned the rule-switch, a similar fraction of cells showed selectivity for the new categories, whereas selectivity for the old, now irrelevant categories ceased (rule 1:  $0.07\% \pm 0.05\%$ , rule 2:  $8.6\% \pm 2.8\%$ ) (Fig. [2h](#), Extended Data Fig. [6a](#)).

To convert an internal category representation into a motor decision, it would be sufficient for cells in the mPFC to show selectivity for only one category<sup>31</sup>. However, we observed two types of neuron—one that represented rewarded stimuli (Go preferring: 73% of all category-selective cells at T5 and 65% at T8) and the other non-rewarded stimuli (NoGo preferring: 27% at T5 and 35% at T8). Thus, cells in the mouse mPFC develop flexible representations of rule-based categories over the course of learning.

## Category selectivity emerges over time

Our chronic recording approach allowed us to ask whether the cells that coded for learned categories in rule 2 were the same ones that had represented categories in rule 1. Although many cells that were category-selective for rule 1 were less selective for rule 2, a subset of neurons remained category-selective throughout (Fig. [3a](#), Extended Data Fig. [7a,b](#)). We found that, on average, the Go category-selective neurons remapped their responses to the new Go category—that is, after the rule-switch, they responded to a different set of visual stimuli. By contrast, the NoGo category-selective cells did not remap (Fig. [3a](#), Extended Data Fig. [7c,d](#)). They lost their selectivity after

the rule-switch, and a new set of cells became NoGo category-selective for the newly defined categories. Similarly, the Go category-selective cells observed after the rule-switch showed previous selectivity to the first rule, whereas rule 2 NoGo category-selective neurons did not show any selectivity before the rule-switch on average (Fig. 3b). In line with this, we observed that the Go category-selective populations for each rule overlapped more than expected by chance (Methods), in contrast to NoGo category-selective populations (Extended Data Fig. 6b–d). Notably, neurons were less likely than chance to switch their preference from Go to NoGo and vice versa (Extended Data Fig. 6b).

**Fig. 3: Two populations of category-selective neurons show different dynamics during a rule-switch.**

 figure3

**a**, Left, CTI of all category-selective neurons identified at T5 (grey highlight), shown for time points T1, T5 and T8.  $P_{T1-T5} = 1.1 \times 10^{-36}$ ,  $P_{T1-T8} = 1.6 \times 10^{-14}$ ,  $P_{T5-T8} = 6.9 \times 10^{-27}$ , two-tailed WMPSR test, Bonferroni-corrected for three comparisons ( $n = 213$  cells). Black line denotes the mean. Right, average inferred spike rate per stimulus of Go and NoGo category-selective cells at T1, T5 and T8 ( $n_{Go} = 156$  cells;  $n_{NoGo} = 57$  cells) **b**, As in **a**, but for category-selective cells defined at T8.  $P_{T1-T5} = 4.2 \times 10^{-18}$ ,  $P_{T1-T8} = 2.9 \times 10^{-33}$ ,  $P_{T5-T8} = 1.1 \times 10^{-27}$ , two-tailed WMPSR test, Bonferroni-corrected for three comparisons ( $n = 192$  cells;  $n_{Go} = 122$  cells;  $n_{NoGo} = 70$  cells). **c**, Left, inferred spike rate of all Go (top) and NoGo (bottom) category-selective

cells, identified at T5 (grey highlight), during trials of all Go (green) and NoGo (red) category stimuli at T1–T8. Grey denotes stimulus presentation. Data are mean  $\pm$  s.e.m., across cells. Right, inferred spike rate during stimulus presentation of all Go (top) and NoGo (bottom) category-selective cells. Green denotes Go category, red denotes NoGo category, orange area denotes the difference. Black denotes the mean inferred spike rate in the pre-stimulus period. Data are mean  $\pm$  s.e.m., across cells. **d**, As in **c**, for category-selective cells defined at T8.

[Source data](#)

[Full size image](#)

It is currently debated whether such flexible representations in the PFC are gradually built up during learning—that is, are part of the memory of the learned categories—or whether they are instantaneously assigned during the task to represent anything that becomes relevant to the animal<sup>7,8,9</sup>. This question can be answered only by monitoring neurons throughout the learning process, starting from a naive animal. We took advantage of the fact that our mice had never been trained on a categorization task and we followed the development of category-selective responses of individual neurons over the entire time course of rule-based category learning (Fig. 3c). Focusing on the period over which selectivity emerged, we observed a marked difference between the time courses that the Go and NoGo category-selective neurons followed. On average, the Go category-selective cells showed large, stable responses for the Go category, early on after presentation of the initial category stimuli in an ad hoc fashion (T2–T5) (Fig. 3c, Extended Data Fig. 7e). By contrast, the NoGo category-selective cells only gradually developed selectivity with increasing categorization demand of the task (T4, T5) (Fig. 3c, Extended Data Fig. 7f). After the rule-switch, the Go category-selective cells on average switched their stimulus selectivity, thereby retaining category selectivity. Former NoGo category-selective cells gradually lost selectivity, whereas a new, independent population of NoGo category-selective neurons gained selectivity (Fig. 3c, d). Notably, after the rule-switch, Go category-selective neurons showed increased Go responsiveness beyond a stable level of Go selectivity during earlier training (Fig. 3d).

A possible explanation for the different time courses could be that various task-relevant components differentially contribute to the average selectivity. It is well established that—beyond the category selectivity we observed—the mPFC contains representations of choice and reward<sup>32,33,34,35</sup>. In our paradigm, choice and reward associations are learned earlier than categories, and stay constant through the rule-switch. Therefore, neurons selective for choice and reward are expected to show a different time course than neurons selective for stimulus category (Extended Data Fig. 7g). We identified individual neurons that acquire selectivity early-on during task learning as well as neurons that develop selectivity more gradually, with increasing

categorization demand (Extended Data Fig. 7*h–k*). In line with their average (Fig. 3*c*, *d*), most NoGo-preferring neurons followed the gradual time course, reflecting acquisition of the respective category rule, whereas Go-preferring neurons followed either of the time courses (Extended Data Fig. 7*h–k*). Thus, neurons that prefer the Go category were modulated by category, as well as by the earlier learned reward and choice associations (Extended Data Fig. 8).

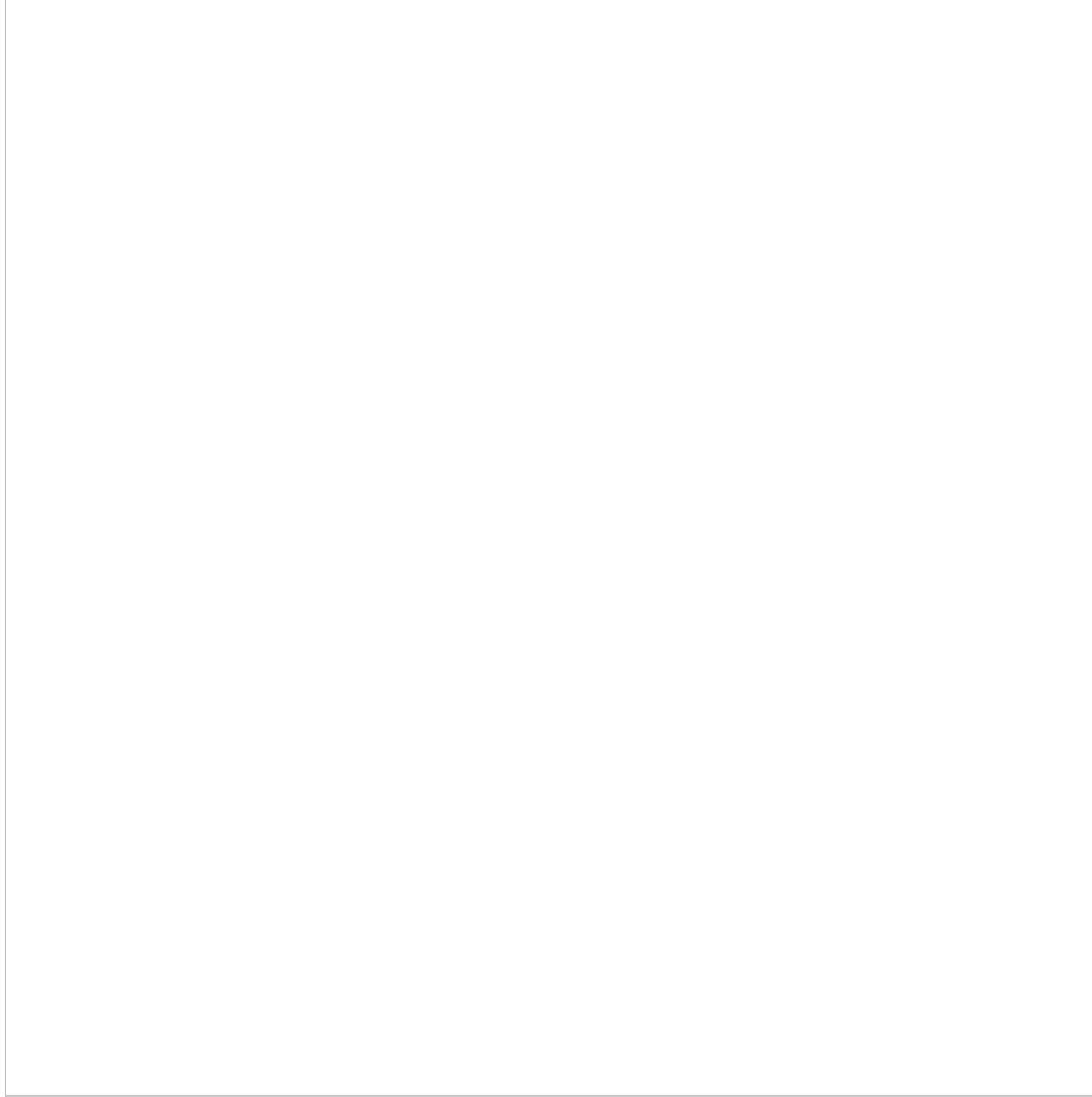
To disentangle how stimulus category, choice and reward affected the trial-by-trial responses of category-selective neurons, we used linear regression to determine their individual contributions (Extended Data Fig. 9*a*). Although choice selectivity did not directly explain CTI (Extended Data Fig. 9*b,c*), the activity pattern of Go category-selective cells showed significant modulation by multiple factors, stimulus category, choice and reward (Extended Data Fig. 9*d*). By contrast, the responses of NoGo category-selective cells were only significantly influenced by category identity (Extended Data Fig. 9*d*). We performed hierarchical clustering to explore the entire task-responsive neuronal population in the mPFC including category-selective cells and found clusters of mPFC neurons that were predominantly modulated by a single parameter—that is, category, choice (lick) and reward (Extended Data Fig. 9*e–i*, cluster number 1, 2 and 3, respectively). In addition, there were also clusters of neurons modulated by specific combinations of task parameters (Extended Data Fig. 9*i*, clusters number 4, 5 and 9). These results are in line with recent studies in primates and mice, reporting mixed selectivity of neurons in the PFC after animals learned cognitive tasks<sup>21,36,37,38</sup>. In summary, the mouse mPFC contains neurons modulated by a single parameter (such as category) and neurons that show mixed-selective responses.

## Category tuning generalizes across tasks

Because the activity of many mPFC neurons, including category-selective neurons, correlated with combinations of stimulus category, choice and reward, we aimed to experimentally determine the unique category-selective component. Exclusively category-modulated neurons can be revealed by experimental decoupling of the presented category and the associated motor response<sup>39</sup>. Within the framework of our rule-based categorization paradigm, we achieved this by initially training mice to categorize in the Go/NoGo task (as before), and then changing the task to a left/right choice design (Fig. 4*a*). As a consequence, the previous Go (lick) category changed into a ‘GoRight’ (lick right) category, and the previous NoGo (no lick) category was now also rewarded if the mouse made a ‘GoLeft’ (lick left) response. In this experiment, neurons that were category-selective in the Go/NoGo task could either retain their category selectivity in the left/right choice task (indicating that they are genuinely category-selective), or change their response pattern, reflecting selectivity rather for motor planning, choice or associated reward (Fig. 4*b*).

**Fig. 4: Mouse mPFC contains uniquely category-modulated neurons.**

 figure4



**a**, Schematic depicting the change in task from Go/NoGo (T5) to left/right choice (L/R). The motor response changed from Go to GoRight, and from NoGo to GoLeft. The category rule remained the same. **b**, Possible changes in neuronal responses between T5 and left/right choice. Top, choice/reward-selective neurons. Bottom, uniquely category-selective neurons. **c, d**, throughout category learning and the change in task, aligned to criterion ( $>66\%$  correct,  $n = 9$  mice). **d**, Example HLS maps before (T5) and after (L/R) the task change. Scale bars, 30  $\mu\text{m}$ . Hue: preferred category; lightness: response amplitude; saturation: category selectivity. **e**, CTI of all recorded neurons, calculated using either the relevant or the irrelevant rule, before (T5) and

after (L/R) the task change. T5:  $P = 1.3 \times 10^{-161}$ , L/R:  $P = 1.3 \times 10^{-24}$ , two-tailed WMPSR test ( $n = 2,389$ ). Grey lines denote CTI = 0.1. **f**, Top, inferred spike rate for stimuli ordered along the relevant dimension (black), or the irrelevant dimension (blue) across all Go category-selective neurons (defined at T5, CTI > 0.1).  $P_{T5\ rel} = 6.5 \times 10^{-28}$ ,  $P_{T5\ irrel} = 1.6 \times 10^{-5}$ ,  $P_{L/R\ rel} = 1.6 \times 10^{-19}$ ,  $P_{L/R\ irrel} = 0.38$ , two-tailed WMPSR test ( $n = 407$ ). Data are mean ± s.e.m. Bottom, mean activity per stimulus. **g**, As in **f**, for NoGo category-selective neurons.  $P_{T5\ rel} = 1.9 \times 10^{-4}$ ,  $P_{T5\ irrel} = 0.45$ ,  $P_{L/R\ rel} = 7.7 \times 10^{-4}$ ,  $P_{L/R\ irrel} = 0.21$ , two-tailed WMPSR ( $n = 48$ ). **h**, Predictor weights and response amplitudes of significantly modulated neurons.  $P < 0.05$ , at least one predictor,  $F$ -statistic (1,904 neurons). Scale bars, 50 neurons. Top row, normalized predictor weights for each neuron. Left, neurons with a negative category weight (category 1, NoGo/GoLeft). Middle, neurons with a positive category weight (category 2, Go/GoRight). Right, no significant category modulation. Middle row, average (normalized) activity in correct, incorrect and missed trials of categories 1 and 2 separately. Bottom row, per group, the mean normalized response to all presented stimuli at T5 (left) and L/R (right).

[Source data](#)

[Full size image](#)

We first trained nine mice to categorize visual stimuli according to either the spatial frequency or the orientation rule (the task was identical to that in Fig. 1 and Extended Data Fig. 1, up to the generalization test T5; Fig. 4c). After session T5, we changed the behavioural setup by replacing the single centred lick spout with two laterally placed lick spouts (left/right choice paradigm). The mice quickly adapted to the change and within the first four trials also responded with licks to the previous NoGo category (now GoLeft; note that the ratio between the left and right licks varied throughout the session). Although the mice did not specifically target their first licks to the correct spout, they performed a similar number of licks on both lick spouts and obtained a similar amount of rewards for both categories.

We found a significant proportion of category-selective neurons before (T5) and after the task change (left/right; threshold of CTI > 0.1 according to the relevant rule) (Fig. 4d, e). On average, category-selective cells identified at T5 discriminated the stimulus categories also after the task change (Fig. 4d, f, g, Extended Data Fig. 10a–g), although their selectivity decreased. The left/right choice task allowed us to compare trials with different stimulus categories in the absence of choice and reward (missed trials). Neurons that were initially selective for the Go category remained selective for the same stimulus category. Likewise, initial NoGo category-selective neurons, remained only responsive to stimuli of the previous NoGo category (Extended Data Fig. 10h, i).

However, the overall decrease in selectivity after the task change indicated that also choice-and reward-selective neurons were identified as ‘category-selective’ in the Go/NoGo task (Fig. 4b). Because the left/right choice task changed how reward and motor contingencies mapped onto the stimulus space, but did not change the mapping of category identity, we were able to use a regression model to disambiguate these contributions. Only neurons that remained category-selective across the task change will be significantly fitted by the Category predictor. Apparent category-selective neurons—that is, choice- and reward-modulated neurons, will be better predicted by the Go and Reward predictors. This analysis showed that mouse mPFC neurons represent categories in conjunction with reward and choice. Most importantly, it also revealed a set of uniquely category-modulated neurons in the mPFC (4.3%) (Fig. 4h, Extended Data Fig. 10j).

Recent work has shown the influence of uninstructed behaviours, such as whisking and eye movements, on neuronal response variability in operant tasks<sup>40</sup>. If such behaviours correlated with the category identity of the presented stimuli, they could lead to apparent category selectivity. To control for this, we tracked key postural markers using DeepLabCut<sup>41,42</sup> and combined them with in-task recorded instructed behaviours and task parameters to predict neural activity. We found that there was a significant and unique contribution for all instructed and uninstructed behavioural variables. Notably, however, there was also a unique contribution of the category component that could not be accounted for by any of the instructed or uninstructed behavioural parameters (Extended Data Fig. 10k–o, Supplementary Video 1). We therefore conclude that the mPFC contains a sparse but distinct set of neurons that represent learned categories irrespective of associated motor behaviours and reward.

## Discussion

Using a paradigm to study learning of rule-based categories in mice, we could follow neuronal populations in the mPFC throughout the entire learning process, from naive to expert mice. We found two distinct groups of cells developing a representation of learned categories with different learning-related dynamics. The NoGo category representation emerged gradually, was rule-specific and was not strongly modulated by additional task parameters, in contrast to the Go representation. In addition, we observed that selectivity for the Go category increased further in the fast rule-switch phase compared to the slow, initial learning phase. This difference could be a consequence of Go category-selective neurons belonging to intrinsically different representations of choice, reward and categories. By experimentally decoupling these, we confirmed that many category-selective neurons were actually mixed-selective, which could benefit the representation of task-relevant information<sup>38</sup>. However, the experiment also revealed uniquely category-selective mPFC neurons, for both learned categories. In line with previous studies<sup>3,19,32,43</sup>, we found that the mPFC initially

contains a conjunctive stimulus and choice representation. This representation flexibly followed the novel Go category when a mouse learned the second rule. In parallel, a slowly learning group of Go category-selective cells emerged for each rule, following a time course similar to the NoGo category representation.

This mouse model of rule-based category learning opens up possibilities to causally investigate neuronal interactions across several cortical and subcortical circuits. Many brain areas, such as posterior parietal cortex<sup>4,44</sup>, sensory areas (P.M.G., S.R., T.B. and M.H., manuscript submitted)<sup>44,45</sup> and striatum<sup>3</sup>, contribute to multiple aspects of category learning and categorization behaviour. Several circuit models on areal interactions have been put forward<sup>43,46</sup>. One model of particular interest proposed that slow-learning PFC circuits acquire category selectivity using rapidly learned stimulus-specific activity originating in the striatum as a teaching signal<sup>3,46</sup>. Within this framework, the mPFC could compute the rule-dependent NoGo category representation from the fast-arising activity of conjunctive Go/choice-selective neurons mediated by local inhibitory circuits. Rule-based category learning in mice allows for testing of specific predictions of such circuit models for prefrontal cortex function by observing initially naive mice throughout the learning process. In particular, the possibility to investigate and observe neuronal responses in the mPFC during category learning in mice opens a window to study the neural circuitry that underlies categorization and storage of semantic memories<sup>47</sup> also in this species.

## Methods

### Data reporting

No statistical methods were used to predetermine sample size. Mice were randomly assigned to the categorization rule ‘spatial frequency’ or ‘orientation’. The investigators were not blinded to allocation during experiments and outcome assessment.

### Animals

All procedures were performed in accordance with the institutional guidelines of the Max Planck Society and the local government (Regierung von Oberbayern). Twenty female C57BL/6 mice (postnatal day (P) 63–P82 at the day of surgery) were housed in groups of four to six littermates in standard individually ventilated cages (IVC, Tecniplast GR900). Mice had access to a running wheel and other enrichment material such as a tunnel and a house. All mice were kept on an inverted 12 h light/12 h dark cycle with lights on at 22:00. Before and during the experiment, the mice had ad libitum access to standard chow (1310, Altromin Spezialfutter). Before the start of

behavioural experiments, mice had ad libitum access to water. At the end of the experiments, mice were perfused with 4% paraformaldehyde (PFA) in PBS and their brains were post-fixed in 4% PFA in PBS at 4 °C.

## Surgical procedures

Before surgery, a prism implant was prepared by attaching a 1.5 mm × 1.5 mm prism (aluminium coating on the long side, MPCH-1.5, IMM photonics) to a 0.13 mm thick, 3 mm diameter glass coverslip (41001103, Glaswarenfabrik Karl Hecht) using UV-curing optical glue (Norland optical adhesive 71, Norland Products) and was left to fully cure at room temperature for a minimum of 24 h. Mice were anaesthetized with a mixture of fentanyl, midazolam and medetomidine in saline (0.05 mg kg<sup>-1</sup>, 5 mg kg<sup>-1</sup> and 0.5 mg kg<sup>-1</sup> respectively, injected intraperitoneally). As soon as sufficient depth of anaesthesia was confirmed by absence of the pedal reflex, carprofen in saline (5 mg kg<sup>-1</sup>, injected subcutaneously) was administered for general analgesia. The eyes were covered with ophthalmic ointment (IsoptoMax/Bepanthen) and lidocaine (Aspen Pharma) was applied on and underneath the scalp for topical analgesia. The scull was exposed, dried and subsequently scraped with a scalpel to improve adherence of the head plate. The scalp surrounding the exposed area was adhered to the skull using Histoacryl (B. Braun Surgical). A custom-designed head plate was centred at ML 0 mm, approximately 3 mm posterior to bregma, attached with cyanoacrylate glue (Ultra Gel Matic, Pattex) and secured with dental acrylic (Paladur). A 3 mm diameter craniotomy, centred at anterior–posterior (AP) 1.9 mm, medial–lateral (ML) 0 mm, was performed using a dental drill. The hemisphere for prism insertion was selected based on the pattern of bridging veins. Before inserting the prism, two injections (50 nl min<sup>-1</sup>) of 200–250 nl of virus solution

(AAV2/1.hSyn.mRuby2.GSG.P2A.GCaMP6m.WPRE.SV40, titre:  $1.02 \times 10^{13}$  genome copies (GC) ml<sup>-1</sup>, Plasmid catalogue 51473, Addgene) were targeted at the medial prefrontal cortex opposite to the prism implant, coordinates: AP 1.4 mm to AP 2.8 mm, ML 0.25 mm, dorsal–ventral (DV) 2.3 mm (Nanoject, Neurostar). The left hemisphere was injected in 11 mice, and the right hemisphere in 9 mice. Subsequently, a durotomy was performed using microscissors (15070-08, Fine Science Tools) over the contralateral hemisphere, next to the medial sinus. The prism implant was inserted, gently pushing the medial sinus aside until the target cortical region became visible through the prism (for a detailed description, see ref. [29](#)). The coverslip was attached to the surrounding skull using cyanoacrylate glue and dental acrylic. After surgery, the anaesthesia was antagonized with a mixture of naloxone, flumazenil and atipamezole in saline (1.2 mg kg<sup>-1</sup>, 0.5 mg kg<sup>-1</sup> and 2.5 mg kg<sup>-1</sup> respectively, injected subcutaneously) and the mice were placed under a heat lamp for recovery. Post-operative analgesia was provided for two subsequent days with carprofen (5 mg kg<sup>-1</sup>, injected subcutaneously).

## Visual stimuli

Stimuli for behavioural training were presented in the centre of a gamma corrected LCD monitor (Dell P2414H; resolution: 1,920 by 1,080 pixels; width: 52.8 cm; height: 29.6 cm; maximum luminance: 182.3 Cd m<sup>-2</sup>). The centre of the monitor was positioned at about 0° azimuth and 0° elevation at a distance of 18 cm, facing the mouse straight on. The stimuli were 36 different sinusoidal gratings, each with a specific orientation and spatial frequency combination, shown in full contrast on a grey background (see Extended Data Fig. 1 for schematic of stimuli and task stages). Orientations ranged from 0° to 90°, the spatial frequencies from 0.023 cycles per degrees (cyc/°) to 0.25 cyc/° (orientations: [0, 15, 30, 60, 75, 90] °, spatial frequencies: [0.023, 0.027, 0.033, 0.06, 0.1, 0.25] cyc/°). The stimulus size was 45 retinal degrees in diameter, including an annulus of 4 degrees blending into the equiluminant grey background. The gratings drifted with a temporal frequency of 1.5 cycles per s.

In a subset of experiments ( $n = 3$  mice), a dense stimulus space was presented, consisting of 49 stimuli ranging from 15° to 75° in orientation and from 0.027 cyc/° to 0.1 cyc/° in spatial frequency (orientations: [15, 30, 37.5, 45, 52.5, 60, 75] °, spatial frequencies: [0.027, 0.033, 0.036, 0.043, 0.052, 0.06, 0.1] cyc/°). Stimuli on the category boundary (either having an orientation of 45° or a spatial frequency of 0.043 cyc/°) were assigned to both categories, hence rewarded in 50% of trials.

All stimuli were created and presented using the Psychophysics Toolbox extensions of MATLAB<sup>48,49,50</sup>.

## Behaviour

Behavioural experiments started seven days after surgery. The water restriction regime and the behavioural apparatus were previously described<sup>51</sup>. In short, mice were restricted to 85% of their initial weight on the starting date by individually adjusting the daily water ration. First, mice were accustomed to the experimenter and head fixation in the setup by daily handling sessions lasting 10 min. During these sessions, the water ration was offered in a handheld syringe. The remainder was supplemented in an individual drinking cage after a delay of approximately 30 min. After four to seven days of handling, mice were pre-trained to lick for reward, while being head-fixed on the spherical treadmill<sup>52,53,54</sup> in absence of visual stimulation. Whenever a mouse ceased to run (velocity below 1 cm s<sup>-1</sup>) and made a lick on the spout, a water reward (drop size 8 µl) was delivered via the spout. A baseline imaging time point (T1) was acquired once the mice consumed more than 50 drops per session (35 to 45 min) on two consecutive days (requiring about three days of pre-training).

Subsequently, daily sessions of visual discrimination training for two initial stimuli started. Each mouse was randomly assigned to one of two groups. One group was first trained on the orientation rule, then on the spatial frequency rule. For the other group, the sequence of the rules was reversed (Extended Data Fig. 1). Each rule defined a Go category and a NoGo category, separated by a boundary at either 45° (orientation rule) or at 0.043 cyc/° (spatial frequency rule). Trials started with an inter-trial interval of 5 s. After that, the mouse could initiate stimulus presentation by halting and refraining from licking for a minimum of 0.5 s. A single stimulus was subsequently shown for  $1.3 \pm 0.2$  s. At any time during stimulus presentation, the mouse could make a lick to indicate a Go choice. Trials with a Go choice in response to a Go category stimulus triggered a water reward and were classified as hits; trials in which the mice failed to lick during Go category stimulus presentation were considered misses. Correct withholding of a lick to a NoGo category stimulus was classified as a correct rejection, and did not result in a water reward. A lick during a NoGo category stimulus counted as a false alarm. Initially, false alarms only led to the termination of the current trial; later during training, false alarms were followed by a time-out of 5–7 s showing a time-out stimulus (a narrow, horizontal, black bar). Time-outs were included to reduce a Go bias that mice typically showed. The second imaging session (T2) was carried out after a mouse performed at more than 66% correct Go choices in a given session (requiring 11 to 40 sessions).

For the next training stage (leading up to imaging session T3) further stimuli were added (Extended Data Fig. 1a), such that both the Go category and the NoGo category consisted of three stimuli differing in the feature either irrelevant to the category rule (T3a,  $n = 6$  mice), or relevant to the category rule (T3b,  $n = 5$  mice). Whenever a mouse's performance exceeded 66% correct Go choices in one session, we proceeded to the next training (and imaging) stage; 6 stimuli per category, 9 stimuli per category (imaging session T4), and finally 18 stimuli per category (imaging session T5), the latter serving as a crucial test for generalization behaviour.

**Rule-switch:** After successful learning of rule 1, mice ( $n = 11$ ) were retrained using the previously irrelevant dimension. This stage, known as rule-switch training, started with two exemplar stimuli for the new rule, and then proceeded with the same steps as for rule 1 and ended with another generalization test of rule 2 (18 stimuli per category, imaging session T8).

**Task change:** After successful learning of rule 1 (T5), the categorization performance of mice ( $n = 9$ ) was tested with a different operant response, in a left/right choice task. For this session, the behavioural setup was slightly modified to create a left/right choice task. Instead of one lick spout centred in front of the mouse, the mouse was now presented with two lick-spouts, one offset to the left and one offset to the right. Stimuli of the previous Go category were assigned a new GoRight response (rewarded after a lick on the right lick spout). Stimuli of the previous NoGo category were

assigned a new GoLeft response (rewarded after a lick on the left lick spout). The original stimulus to category assignment—that is, the categorization rule—remained the same throughout the task change. Before the first stimulus presentation, ten drops were manually given on each lick spout to motivate the mice to lick on both sides.

Throughout training, stimuli from the Go category and the NoGo category were presented in a pseudorandomized fashion, showing not more than three stimuli of the same category in a row. The behavioural training program was a custom written MATLAB routine (Mathworks).

## Imaging

Two-photon imaging<sup>55</sup> through the implanted prism was performed at 5–8 time points in each mouse throughout the training paradigm (T3 was omitted in two mice; for detailed timing of the imaging sessions see Extended Data Fig. 1a). In some mice ( $n = 5$ ) we followed two regions in the same mouse; in these cases, two imaging sessions were acquired on consecutive days during the same training stage. Imaging was done using a custom-built two-photon laser-scanning microscope (resonant scanning system) and a Mai Tai eHP Ti:Sapphire laser (Spectra-Physics) tuned to a wavelength of 940 nm. Images were acquired with a sampling frequency of 10 Hz and  $750 \times 800$  pixels per frame. The mice in the task change experiment were imaged using a customized commercially available two-photon laser-scanning microscope (Thorlabs; same laser specifications as described above), operated with Scanimage 4<sup>56</sup>. In these experiments, images were acquired at 30 Hz and  $512 \times 512$  pixels per frame. The average laser power under the objective ranged from 50 to 80 mW. Note that the laser power was higher than for imaging through a conventional cranial window due to a substantial power loss over the prism<sup>29</sup>. We used a  $16\times$ , 0.8 NA, water immersion objective (Nikon) and diluted ultrasound gel (Dahlhausen) on top of the implant as immersion medium. Two photomultiplier tubes detected the red fluorescence signal of the structural protein mRuby2 (570–690 nm) and the green fluorescence signal of GCaMP6m (500–550 nm)<sup>57</sup>. During imaging, the monitor used for stimulus presentation was shuttered to minimize light contamination<sup>58</sup>. The imaging data were acquired using custom LABVIEW software (National Instruments; software modified from the colibri package by C. Seebacher) and the synchronization of imaging data with behavioural readout and stimulus presentation was done using DAQ cards (National Instruments).

## Tracking of postural markers

In two-photon imaging sessions of a subset of experiments, the mouse was video-tracked using infrared cameras (The Imaging Source Europe). Two cameras were aimed at the eyes, and a third camera was positioned at a slight angle behind the

mouse, in order to record body movements in-task. The eyes of the mouse were back-lit by the infrared two-photon imaging laser and the body was illuminated using an infrared light source (740 nm; Thorlabs). Key eye and body features (see Extended Data Fig. 10) were manually defined and automatically annotated using DeepLabCut<sup>41,42</sup>. From the  $x$  and  $y$  coordinates of these features, we calculated three eye parameters and four postural parameters (pupil diameter, eye position, eyelid opening, front paw angle, hind paw angle of the left hind paw, body elongation/rotation, tail angle; see Extended Data Fig. 10). Supplementary Video 1 shows both eye and body cameras of an example mouse.

## Data analysis

The analysis of behaviour and imaging data was performed using custom written MATLAB routines.

### Behavioural data

Behavioural performance is shown as the sensitivity index,  $d'$ . For every training session,  $d'$  was calculated as the difference between the  $z$ -scored hit rate and the  $z$ -scored false alarm rate. The hit rate was defined as the number of correct category 2 trials divided by the total number of category 2 trials per session. Similarly, the false alarm rate was calculated as the number of incorrect category 1 trials divided by the total number of category 1 trials. In case a mouse performed two training sessions at time points T1, T3, T4, T5, T7 and T8, because two regions were imaged, the displayed value in the learning curve is the average across the two imaging sessions.

The fraction of correct Go choices was calculated as the number of hit trials divided by the number of all trials in which the mouse made a Go choice (the sum of ‘hits’ and ‘false alarms’). The number of days until a mouse reached performance criterion was the amount of daily sessions until the fraction of correct Go choices exceeded 0.66. Pre-training sessions without visual stimulation were not included in this quantification.

To investigate categorization behaviour across the entire stimulus space, we calculated the ‘fraction chosen’: The number of Go choices in response to a specific stimulus divided by the total number of presentations for that stimulus (see example in Fig. 1d; for all mice see Extended Data Fig. 2). Finally, we constructed psychometric curves showing the effect of each feature (that is, rule-relevant versus rule-irrelevant) on the behaviour of the mice (Fig. 1j). For that, the stimulus-specific ‘fraction chosen’ values were averaged along the irrelevant or the relevant feature dimension, respectively (see Fig. 1i).

To estimate learning rates, each individual learning curve was fitted with a sigmoid function:

$$y(x) = p_1 + \frac{p_2}{1 + e^{-p_3(x - p_4)}}$$

in which  $p_1$  determines the minimum of the sigmoid curve (for curve fitting fixed to 0),  $p_2$  the maximum,  $p_3$  the slope and  $p_4$  the inflection point. The parameter defining the minimum was fixed at a  $d'$  of 0. Learning curves for rule 1 and rule 2 were fitted independently. Goodness of fit was determined as the root-mean-square error between the learning curve and the fitted curve.

## Imaging data processing

The imaging data were first preprocessed by performing dark-current subtraction (using the average signal intensity during a laser-off period) and line shift correction. Rigid  $xy$  image displacement was first calculated on the structural red fluorescence channel using the cross correlation of the 2D Fourier transform of the images<sup>59</sup>, and subsequently corrected on both channels. For each imaging session, cells were manually segmented using the average image of the red fluorescence channel across the entire session. The cell identity was then manually matched across all imaging time points and only cells that could be identified in every session from T1 to T8 or T5 to left/right were included in the analysis. This criterion excluded one mouse (M06) from all further analyses, because of lost optical access at T8. The average green fluorescence signal was extracted for each cell and then corrected for neuropil contamination by subtracting the signal of 30  $\mu\text{m}$  surrounding each cell multiplied by 0.7 and adding the median multiplied by 0.7 (refs. [57,60](#)). From this fluorescence trace, we calculated  $\Delta F/F$  as  $(F - F_0)/F_0$  per frame.  $F_0$  was defined as the 25th percentile of the fluorescence trace in a sliding window of 60 s. From this trace, we inferred the spiking activity of each cell using the constrained foopsi algorithm<sup>61,62,63</sup>. The inferred spike rate during the stimulus presentation period was used for all further calculations and in all figure panels, except for the HLS maps and the left panels of Fig. [2d,e](#), where we display the  $\Delta F/F$  trace for comparison.

To display lick-triggered neuronal activity (Extended Data Fig. [8](#)), we averaged the inferred spike rate centred on the onset of the mouse's lick-bouts. A lick-bout was defined as a sequence of licks, in which the interval between every two consecutive licks did not exceed 500 ms. Thus, a lick was part of a lick-bout if it happened within 500 ms after the previous lick. The onset of each lick-bout was the time of the first lick in the lick-bout.

## Category-tuning index

For every cell, we calculated the CTI as previously described<sup>30</sup>. In short, we quantified the mean inferred spike rate during stimulus presentation for every stimulus. Next, we calculated the mean difference in inferred rate between stimuli of the same category (within), subtracted it from the mean difference between stimuli belonging to the two different categories (across) and normalized by the sum (across + within). This calculation results in an index ranging from -1 to 1, with category-unselective cells showing CTIs close to and below 0 and an ideal category-selective cell having an index of 1. Category-selective cells were defined as cells with a CTI value larger than 0.1. This threshold was chosen based on the distribution of CTIs in the naive population (T1), where individual cells rarely crossed this value. As a control, we used other thresholds (0.07, 0.15 and 0.20) and found no qualitative difference in the results other than that the fraction of category-selective cells scaled.

The fraction of category-selective cells was calculated as the number of neurons above threshold per imaging region, divided by the total number of chronically recorded neurons in that imaging region. Category-selective cells, determined by their CTI at time points T5 and T8, were divided in a Go category-selective and a NoGo category-selective group; neurons with higher average activity in Go category trials than in NoGo category trials were grouped as Go category-selective cells and conversely, cells with a higher average activity in NoGo category trials were labelled as NoGo category-selective. The overlap between the Go and NoGo category-selective groups was calculated between T5 and T8. The expected range of overlap assuming random independent sampling was calculated from the data, but with shuffled neuron identities (using the 95% percentile of the shuffled distribution). For time points at which not all stimuli were presented (T2, T3, T4, T6 and T7), we approximated category-tuning from the average responses to Go category trials and NoGo category trials.

## Bayesian decoding

We decoded category identity from trial-by-trial activity patterns of a single neuron up to groups of ten neurons using Bayes theorem:

$$p(c|r) = \frac{p(r|c)p(c)}{p(r)}$$

in which  $p(r|c)$  is the probability of a single trial response  $r$  when observed in either category 1 or 2 trials (calculated from an exponential distribution),  $p(c)$  as the prior probability of observing each category, and  $p(r)$  as the probability of observing the response. To cross-validate decoding performance, trials were first split into a training and test set (70% and 30%, respectively). The trial-averaged inferred spike rates followed an exponential distribution, which we estimated for each category individually (using the training set). Then, for each trial in the test set, we calculated the probability that the neuronal response came from those distributions. The

distribution that gave the higher probability was determined as the decoder's prediction. Decoder performance was calculated as the fraction of correctly predicted trials. As a control, decoding performance was also calculated after shuffling category identities across trials.

## Selectivity time course

Average selectivity of individual neurons was calculated as the mean difference between responses to all Go category stimuli and all NoGo category stimuli, at every imaging time point (T1–T8). For linear regression, we defined three characteristic selectivity time courses (shown in Extended Data Fig. 2), resembling acquired selectivity for reward/choice, categorization rule 1 and categorization rule 2. Within each of these time courses, maximum selectivity was assigned the value 1 and no selectivity the value 0. The characteristic time courses were used as predictors in a model fitting the development of selectivity of individual neurons over time.

## Generalized linear models to assess the influence of individual task parameters

We performed multilinear regression on neurons that were identified in all imaging time points of the rule-switch experiment. The regression model predicted the trial-wise mean spike rate of each cell during the stimulus presentation periods at imaging time point T5. Categorical predictors were: Category identity of the presented stimulus (0: category 1, 1: category 2), choice of the mouse (0: NoGo, 1: Go), and reward (0: no reward, 1: reward). The average running speed during the trial was modelled as a continuous predictor. A positive predictor weight indicated that the activity of a neuron was increased in trials where the value of the predictor was higher. A negative predictor weight reflected an inverse relation between the predictor's value and the neuron's firing rate. We normalized the predictor weights for overall differences in response amplitudes, by dividing each weight by the sum of all absolute predictor weights (including the intercept).

Hierarchical clustering was performed on relative predictor weights of neurons, including only cells with an  $R^2$  value larger than 0.05. The optimal number of clusters was calculated using gap statistic values, determined as the smallest cluster number  $k$  that fulfilled the criterion (here nine clusters):

$$\$\\{ \\rm{Gap} \\} (k) \\geq \\{ \\rm{Gapmax} \\} - \\{ \\rm{s.e.} \\} \\{ \\rm{e.} \\} / (\\{ \\rm{Gapmax} \\})$$

in which  $\text{Gap}(k)$  is the gap statistic for  $k$  clusters,  $\text{Gapmax}$  is the largest gap value, and  $\text{s.e.}(\text{Gapmax})$  is the standard error corresponding to the largest gap value.

We obtained linkage and relative predictor weights of the clusters from the MATLAB clusterdata algorithm.

To probe the influence of operant motor behaviour in the task change experiment, we concatenated all trials of sessions T5 (generalization session, Go/NoGo task) and L/R (left/right choice task). A stepwise linear regression model predicted the trial-averaged inferred spike rate of all recorded neurons individually. The predictors were the following categorical variables: category identity of the stimulus (0: category 1; 1: category 2), Go response of the mouse (0: NoGo, 1: all forms of Go, that is, Go/GoRight/GoLeft), reward (0: no reward, 1: reward) and two predictors that were specific to a motor response in the left/right session: GoRight and GoLeft. We only considered significant predictor weights, determined from an *F*-statistic comparing a model with and without a predictor. Predictor weights were normalized by dividing each weight by the maximum of all predictor weights.

## Linear regression assessing the influence of instructed and uninstructed behaviours

The trial-averaged inferred spike rate of all recorded neurons in session T5 of a subset of experiments was fitted using a linear model. Body and eye parameters describing uninstructed behaviours were included in the model as continuous predictors. In addition, we included three categorical task-relevant predictors: category identity of the presented stimulus, choice of the mouse, and reward. For each predictor, we determined its maximum predictive power ( $\text{cv}R^2$ ) and its unique contribution ( $\Delta R^2$ ), similar to the approach previously described<sup>40</sup>. Maximum predictive power ( $\text{cv}R^2$ ) was calculated as the predictive performance ( $R^2$ ) of a model with all parameters shuffled, except for the parameter of interest. A parameter's unique contribution ( $\Delta R^2$ ) was quantified as the difference between the full model's  $R^2$  and the  $R^2$  of a model in which the parameter of interest was shuffled.

## Stereotaxic coordinates of imaging regions

We determined the stereotaxic coordinates of the centres of all imaging regions (included in Fig. 2g, h) to place the imaged regions within a common reference frame (Mouse Brain Atlas)<sup>64</sup>. First, we cut 60- $\mu\text{m}$  thick sagittal sections of both hemispheres using a freezing microtome. The AP coordinates outlining the full extent of the prism were identified from a section of the hemisphere into which the prism had been implanted (Extended Data Fig. 4). On the basis of this information, we calculated the exact AP coordinate of the centre of each imaging field of view. We calculated the dorso-ventral coordinate relative to the brain surface, which was aligned with the dorsal border of the prism. Finally, we determined the medio-lateral coordinate of the

imaged field of view from the imaging depth of the field of view relative to the medial pia mater.

## Statistical procedures

All data are presented as mean  $\pm$  s.e.m. unless stated otherwise. Tests for normal distribution were carried out using the Kolmogorov–Smirnov test. Normally distributed data were tested using the two-tailed paired-samples *t*-test. Non-normally distributed data were tested using the two-tailed WMPSR test for paired samples, and the Kruskal–Wallis test for multiple, independent groups. A Bonferroni alpha correction was applied when multiple tests were done on the same data. Correlations were assessed using Pearson’s correlation coefficient, if the data were normally distributed along both axes; otherwise, Spearman’s correlation was applied.

## 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 of this study are available on publication at [https://ging-node.org/sreinert/Category-learning\\_mPFC](https://ging-node.org/sreinert/Category-learning_mPFC). [Source data](#) are provided with this paper.

## Code availability

The custom written MATLAB routines used for data collection and analysis are available upon reasonable request.

## References

1. 1.

Wallis, J. D., Anderson, K. C. & Miller, E. K. Single neurons in prefrontal cortex encode abstract rules. *Nature* **411**, 953–956 (2001).

[ADS](#) [CAS](#) [PubMed](#) [Google Scholar](#)

2. 2.

Freedman, D. J. & Assad, J. A. Experience-dependent representation of visual categories in parietal cortex. *Nature* **443**, 85–88 (2006).

[ADS](#) [CAS](#) [PubMed](#) [Google Scholar](#)

3. 3.

Antzoulatos, E. G. & Miller, E. K. Differences between neural activity in prefrontal cortex and striatum during learning of novel abstract categories. *Neuron* **71**, 243–249 (2011).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

4. 4.

Brincat, S. L., Siegel, M., von Nicolai, C. & Miller, E. K. Gradual progression from sensory to task-related processing in cerebral cortex. *Proc. Natl Acad. Sci. USA* **115**, E7202–E7211 (2018).

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

5. 5.

Goodwin, S. J., Blackman, R. K., Sakellaridi, S. & Chafee, M. V. Executive control over cognition: stronger and earlier rule-based modulation of spatial category signals in prefrontal cortex relative to parietal cortex. *J. Neurosci.* **32**, 3499–3515 (2012).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

6. 6.

Freedman, D. J., Riesenhuber, M., Poggio, T. & Miller, E. K. Categorical representation of visual stimuli in the primate prefrontal cortex. *Science* **291**, 312–316 (2001).

[ADS](#) [CAS](#) [PubMed](#) [Google Scholar](#)

7. 7.

Roy, J. E., Riesenhuber, M., Poggio, T. & Miller, E. K. Prefrontal cortex activity during flexible categorization. *J. Neurosci.* **30**, 8519–8528 (2010).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

8. 8.

Duncan, J. An adaptive coding model of neural function in prefrontal cortex. *Nat. Rev. Neurosci.* **2**, 820–829 (2001).

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

9. 9.

Miller, E. K. & Cohen, J. D. An integrative theory of prefrontal cortex function. *Annu. Rev. Neurosci.* **24**, 167–202 (2001).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

10. 10.

Smith, J. D. et al. Implicit and explicit categorization: a tale of four species. *Neurosci. Biobehav. Rev.* **36**, 2355–2369 (2012).

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

11. 11.

Smith, J. D., Beran, M. J., Crossley, M. J., Boomer, J. & Ashby, F. G. Implicit and explicit category learning by macaques (*Macaca mulatta*) and humans (*Homo sapiens*). *J. Exp. Psychol. Anim. Behav. Process.* **36**, 54–65 (2010).

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

12. 12.

Shepard, R. N. & Chang, J.-J. Stimulus generalization in the learning of classifications. *J. Exp. Psychol.* **65**, 94–102 (1963).

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

13. 13.

Ashby, F. G. & Maddox, W. T. Human category learning. *Annu. Rev. Psychol.* **56**, 149–178 (2005).

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

14. 14.

Ashby, F. G. & Spiering, B. J. The neurobiology of category learning. *Behav. Cogn. Neurosci. Rev.* **3**, 101–113 (2004).

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

15. 15.

Smith, J. D. et al. Pigeons' categorization may be exclusively nonanalytic. *Psychon. Bull. Rev.* **18**, 414–421 (2011).

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

16. 16.

Broschard, M. B., Kim, J., Love, B. C., Wasserman, E. A. & Freeman, J. H. Selective attention in rat visual category learning. *Learn. Mem.* **26**, 84–92 (2019).

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

17. 17.

Jiang, X. et al. Categorization training results in shape- and category-selective human neural plasticity. *Neuron* **53**, 891–903 (2007).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

18. 18.

Ragazzino, M. E., Detrick, S. & Kesner, R. P. Involvement of the prelimbic-infralimbic areas of the rodent prefrontal cortex in behavioral flexibility for place and response learning. *J. Neurosci.* **19**, 4585–4594 (1999).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

19. 19.

Rikhye, R. V., Gilra, A. & Halassa, M. M. Thalamic regulation of switching between cortical representations enables cognitive flexibility. *Nat. Neurosci.* **21**, 1753–1763 (2018).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

20. 20.

de Bruin, J. P., Sánchez-Santed, F., Heinsbroek, R. P., Donker, A. & Postmes, P. A behavioural analysis of rats with damage to the medial prefrontal cortex using the Morris water maze: evidence for behavioural flexibility, but not for impaired spatial navigation. *Brain Res.* **652**, 323–333 (1994).

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

21. 21.

Mansouri, F. A., Matsumoto, K. & Tanaka, K. Prefrontal cell activities related to monkeys' success and failure in adapting to rule changes in a Wisconsin Card Sorting Test analog. *J. Neurosci.* **26**, 2745–2756 (2006).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

22. 22.

Brigman, J. L. & Rothblat, L. A. Stimulus specific deficit on visual reversal learning after lesions of medial prefrontal cortex in the mouse. *Behav. Brain Res.* **187**, 405–410 (2008).

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

23. 23.

Uylings, H. & van Eden, C. G. Chapter 3 Qualitative and quantitative comparison of the prefrontal cortex in rat and in primates, including humans. *Prog. Brain Res.* **85**, 31–62 (1991).

[Google Scholar](#)

24. 24.

Chang, J.-Y., Chen, L., Luo, F., Shi, L.-H. & Woodward, D. J. Neuronal responses in the frontal cortico-basal ganglia system during delayed matching-to-sample task: ensemble recording in freely moving rats. *Exp. Brain Res.* **142**, 67–80 (2002).

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

25. 25.

Kesner, R. P. Subregional analysis of mnemonic functions of the prefrontal cortex in the rat. *Psychobiology (Austin Tex.)* **28**, 219–228 (2000).

[Google Scholar](#)

26. 26.

Seamans, J. K., Lapish, C. C. & Durstewitz, D. Comparing the prefrontal cortex of rats and primates: insights from electrophysiology. *Neurotox. Res.* **14**, 249–262 (2008).

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

27. 27.

Carlén, M. What constitutes the prefrontal cortex? *Science* **358**, 478–482 (2017).

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

28. 28.

Laubach, M., Amarante, L. M., Swanson, K. & White, S. R. What, if anything, is rodent prefrontal cortex? *eNeuro* **5**, ENEURO.0315-18.2018 (2018).

[Google Scholar](#)

29. 29.

Low, R. J., Gu, Y. & Tank, D. W. Cellular resolution optical access to brain regions in fissures: imaging medial prefrontal cortex and grid cells in entorhinal cortex. *Proc. Natl Acad. Sci. USA* **111**, 18739–18744 (2014).

[ADS](#) [CAS](#) [PubMed](#) [Google Scholar](#)

30. 30.

Freedman, D. J., Riesenhuber, M., Poggio, T. & Miller, E. K. Visual categorization and the primate prefrontal cortex: neurophysiology and behavior. *J. Neurophysiol.* **88**, 929–941 (2002).

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

31. 31.

Fitzgerald, J. K. et al. Biased associative representations in parietal cortex. *Neuron* **77**, 180–191 (2013).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

32. 32.

Pinto, L. & Dan, Y. Cell-type-specific activity in prefrontal cortex during goal-directed behavior. *Neuron* **87**, 437–450 (2015).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

33. 33.

Otis, J. M. et al. Prefrontal cortex output circuits guide reward seeking through divergent cue encoding. *Nature* **543**, 103–107 (2017).

[ADS](#) [CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

34. 34.

Huda, R. et al. Distinct prefrontal top-down circuits differentially modulate sensorimotor behavior. *Nat. Commun.* **11**, 6007 (2020).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

35. 35.

Li, B., Nguyen, T. P., Ma, C. & Dan, Y. Inhibition of impulsive action by projection-defined prefrontal pyramidal neurons. *Proc. Natl Acad. Sci. USA* **117**, 17278–17287 (2020).

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

36. 36.

Grunfeld, I. S. & Likhtik, E. Mixed selectivity encoding and action selection in the prefrontal cortex during threat assessment. *Curr. Opin. Neurobiol.* **49**, 108–115 (2018).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

37. 37.

Karlsson, M. P., Tervo, D. G. & Karpova, A. Y. Network resets in medial prefrontal cortex mark the onset of behavioral uncertainty. *Science* **338**, 135–139 (2012).

[ADS](#) [CAS](#) [PubMed](#) [Google Scholar](#)

38. 38.

Rigotti, M. et al. The importance of mixed selectivity in complex cognitive tasks. *Nature* **497**, 585–590 (2013).

[ADS](#) [CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

39. 39.

Freedman, D. J. & Assad, J. A. Neuronal mechanisms of visual categorization: an abstract view on decision making. *Annu. Rev. Neurosci.* **39**, 129–147 (2016).

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

40. 40.

Musall, S., Kaufman, M. T., Juavinett, A. L., Gluf, S. & Churchland, A. K. Single-trial neural dynamics are dominated by richly varied movements. *Nat. Neurosci.* **22**, 1677–1686 (2019).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

41. 41.

Nath, T. et al. Using DeepLabCut for 3D markerless pose estimation across species and behaviors. *Nat. Protocols* **14**, 2152–2176 (2019).

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

42. 42.

Mathis, A. et al. DeepLabCut: markerless pose estimation of user-defined body parts with deep learning. *Nat. Neurosci.* **21**, 1281–1289 (2018).

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

43. 43.

Ashby, F. G., Alfonso-Reese, L. A., Turken, A. U. & Waldron, E. M. A neuropsychological theory of multiple systems in category learning. *Psychol. Rev.* **105**, 442–481 (1998).

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

44. 44.

Zhong, L. et al. Causal contributions of parietal cortex to perceptual decision-making during stimulus categorization. *Nat. Neurosci.* **22**, 963–973 (2019).

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

45. 45.

Xin, Y. et al. Sensory-to-category transformation via dynamic reorganization of ensemble structures in mouse auditory cortex. *Neuron* **103**, 909–921.e6 (2019).

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

46. 46.

Villagrasa, F. et al. On the role of cortex-basal ganglia interactions for category learning: A neuro-computational approach. *J. Neurosci.* **38**, 9551–9562 (2018).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

47. 47.

Tonegawa, S., Morrissey, M. D. & Kitamura, T. The role of engram cells in the systems consolidation of memory. *Nat. Rev. Neurosci.* **19**, 485–498 (2018).

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

48. 48.

Brainard, D. H. The psychophysics toolbox. *Spat. Vis.* **10**, 433–436 (1997).

[CAS](#) [Google Scholar](#)

49. 49.

Pelli, D. G. The VideoToolbox software for visual psychophysics: transforming numbers into movies. *Spat. Vis.* **10**, 437–42 (1997).

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

50. 50.

Kleiner, D. et al. What's new in psychtoolbox-3? *Perception* **36**, 1–16 (2007).

[Google Scholar](#)

51. 51.

Goltstein, P. M., Reinert, S., Glas, A., Bonhoeffer, T. & Hübener, M. Food and water restriction lead to differential learning behaviors in a head-fixed two-choice visual discrimination task for mice. *PLoS ONE* **13**, e0204066 (2018).

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

52. 52.

Hölscher, C., Schnee, A., Dahmen, H., Setia, L. & Mallot, H. A. Rats are able to navigate in virtual environments. *J. Exp. Biol.* **208**, 561–569 (2005).

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

53. 53.

Dombeck, D. A., Khabbaz, A. N., Collman, F., Adelman, T. L. & Tank, D. W. Imaging large-scale neural activity with cellular resolution in awake, mobile mice. *Neuron* **56**, 43–57 (2007).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

54. 54.

Guo, Z. V. et al. Procedures for behavioral experiments in head-fixed mice. *PLoS ONE* **9**, e88678 (2014).

[ADS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

55. 55.

Denk, W., Strickler, J. H. & Webb, W. W. Two-photon laser scanning fluorescence microscopy. *Science* **248**, 73–76 (1990).

[ADS](#) [CAS](#) [PubMed](#) [Google Scholar](#)

56. 56.

Pologruto, T. A., Sabatini, B. L. & Svoboda, K. ScanImage: flexible software for operating laser scanning microscopes. *Biomed. Eng. Online* **2**, 13 (2003).

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

57. 57.

Chen, T. W. et al. Ultrasensitive fluorescent proteins for imaging neuronal activity. *Nature* **499**, 295–300 (2013).

[ADS](#) [CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

58. 58.

Leinweber, M. et al. Two-photon calcium imaging in mice navigating a virtual reality environment. *J. Vis. Exp.* **84**, e50885 (2014).

[Google Scholar](#)

59. 59.

Guizar-Sicairos, M., Thurman, S. T. & Fienup, J. R. Efficient subpixel image registration algorithms. *Opt. Lett.* **33**, 156–158 (2008).

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

60. 60.

Kerlin, A. M., Andermann, M. L., Berezovskii, V. K. & Reid, R. C. Broadly tuned response properties of diverse inhibitory neuron subtypes in mouse visual cortex. *Neuron* **67**, 858–871 (2010).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

61. 61.

Vogelstein, J. T. et al. Fast nonnegative deconvolution for spike train inference from population calcium imaging. *J. Neurophysiol.* **104**, 3691–704 (2010).

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

62. 62.

Pnevmatikakis, E. A. et al. Simultaneous denoising, deconvolution, and demixing of calcium imaging data. *Neuron* **89**, 285–299 (2016).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

63. 63.

Giovannucci, A. et al. CaImAn an open source tool for scalable calcium imaging data analysis. *eLife* **8**, e38173 (2019).

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

64. 64.

Franklin, K. B. & Paxinos, G. *The Mouse Brain in Stereotaxic Coordinates* Vol. 3 (Academic, 2007).

[Download references](#)

## Acknowledgements

We thank C. Leibold for discussions contributing to the study design; V. Staiger, C. Huber, F. Voss and M. Sperling for technical assistance; A. Glas for help with the implant method and discussions; J. Bauer, D. Mearns and T. Mrsic-Flogel for discussions and comments on the manuscript; T. Rose for the viral construct; J. Kuhl for illustrations. This project was funded by the Max Planck Society and the Collaborative Research Center SFB870 (project numbers A07, A08) of the German Research Association (DFG) to T.B. and M.H. and the German Research Association (DFG) via the RTG 2175 ‘Perception in context and its neural basis’ to M.H. and C. Leibold.

## Funding

Open access funding provided by Max Planck Society.

## Author information

### Affiliations

1. Max Planck Institute of Neurobiology, Martinsried, Germany  
Sandra Reinert, Mark Hübener, Tobias Bonhoeffer & Pieter M. Goltstein
2. Graduate School of Systemic Neurosciences, Ludwig-Maximilians-Universität München, Martinsried, Germany

Sandra Reinert

### Authors

1. Sandra Reinert  
[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

2. Mark Hübener

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

3. Tobias Bonhoeffer

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

4. Pieter M. Goltstein

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

## Contributions

S.R., M.H., T.B. and P.M.G. designed the study. S.R. conducted the experiments. S.R. and P.M.G. analysed the data. S.R., M.H., T.B. and P.M.G. wrote the manuscript.

## Corresponding author

Correspondence to [Pieter M. Goltstein](#).

## Ethics declarations

### Competing interests

The authors declare no competing interests.

## Additional information

**Peer review information** *Nature* thanks Anne Churchland and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

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

## Extended data figures and tables

[\*\*Extended Data Fig. 1 Timeline of behavioural training, presented stimuli and learning performance of individual mice for both categorization rules.\*\*](#)

**a**, Timeline showing behavioural training stages, the number of training sessions that mice spent in each stage (min–max) and the imaging sessions (T1–T8). **b**, Stimuli used for category training, aligned to the stages shown in **a**. The scheme shows stimuli for mice that were trained on the spatial frequency rule first, and the orientation rule second. **c**, As in **b**, but for mice trained on the orientation rule first. **d**, Per mouse, the learning curve for training on rule 1. Blue curve denotes single session  $d'$ . Orange curve denotes sigmoid fit of  $d'$ . Arrows indicate imaging time points T2, T3 and T4. **e**, As in **d**, but for rule 2. Arrows indicate imaging time points T6 and

T7. **f**, Parameters describing the fitted sigmoidal curves, comparing rules 1 and 2. Left, maximum.  $P = 0.71$ , two-tailed paired-samples  $t$ -test ( $n = 11$ ). Middle, slope.  $P = 1.9 \times 10^{-5}$ , two-tailed paired-samples  $t$ -test ( $n = 11$ ). Right, inflection point.  $P = 9.3 \times 10^{-6}$ , two-tailed paired-samples  $t$ -test ( $n = 11$ ). **g**, Root-mean-square error (RMSE) of sigmoid fit.  $P = 0.013$ , two-tailed paired-samples  $t$ -test ( $n = 11$ ). **h**,  $d'$  of all mice comparing naive and learned discrimination of the initial two stimuli for the first rule (left) and the second rule (right). Black line indicates the mean across all mice, grey lines represent data of individual mice. [Source data](#)

### **Extended Data Fig. 2 Categorization, generalization and rule-switch performance for individual mice.**

Performance as the fraction of Go choices per stimulus, averaged over the imaging time points for each mouse individually. The time point ‘Learned 2 stim RS’ shows performance after the rule-switch was successfully learned. This time point was not an imaging session. Three mice learned the rule-switch during session T6 (‘single session’). **a**, Mice first trained on the spatial frequency rule and then on the orientation rule (data of M03 is also shown in Fig. [1d,g](#)). **b**, As in **a** for all mice trained initially on the orientation rule and then on the spatial frequency rule. Mouse M06: imaging sessions T7 and T8 were ‘not included’ owing to poor imaging quality; Mice M01 and M02: imaging session T3 was not recorded. [Source data](#)

### **Extended Data Fig. 3 Generalization of stimuli at their first presentation and categorization of stimuli close to, and at the category boundary.**

**a**, Left, schematic of stimulus space during the generalization session (T5). Middle, category choice for every stimulus on its first presentation for an example mouse, green: Go choice, red: NoGo choice. Right, category choice at the first stimulus presentation, averaged across mice ( $n = 10$  mice). **b**,  $d'$  for experienced stimuli and novel stimuli separately, calculated using only the first presentation of each stimulus at the generalization session T5.  $P = 0.19$ , two-tailed paired-samples  $t$ -test ( $n = 10$ ). Grey lines

denote individual mice. Data are mean  $\pm$  s.e.m. (across mice). **c**, Mice use the second rule to categorize stimuli that were only experienced during training on the first rule. Left, schematic showing category identity of stimuli at T5 (Go or NoGo) and whether they were experienced throughout category training on rule 1 (Exp) or novel (Nov). Middle, the highlighted quadrant (green) was part of the Go category; stimuli from this quadrant had been incrementally used throughout category learning up to T5. After T5, mice were trained on the second rule, using only stimuli in the bottom half of the category space (which corresponded to the NoGo category at T5). Right, in the second generalization session (T8, rule-switch generalization), mice were once more exposed to the full category space. Now, the same highlighted quadrant (red) required a NoGo response. However, so far these stimuli were extensively (and only) experienced as requiring a Go response. If mice showed a different fraction of Go choices in T5 and T8, it would reflect category generalization of rule 2, because the absence of experience with the stimuli in the highlighted quadrant prevented learning of a stimulus-response mapping. **d**, Fraction of Go category choices for the first presentation of stimuli highlighted in **c**, comparing T5 and T8. Data are mean  $\pm$  s.e.m. (fraction chosen across all mice). Grey lines indicate data of individual mice, of which some overlap.  $P = 0.002$ , two-tailed paired-samples  $t$ -test ( $n = 10$ ). **e**, Schematic as in **c**, indicating the stimuli and the rule for which  $d'$  was calculated in **f**. **f**,  $d'$  for the first presentation of stimuli highlighted in **c**, comparing T5 and T8. Grey lines denote individual mice.  $P = 2.2 \times 10^{-5}$ , two-tailed paired-samples  $t$ -test ( $n = 10$ ). Data are mean  $\pm$  s.e.m. (across mice). **g**, Schematic indicating the relevant and irrelevant stimulus dimension for the spatial frequency rule (left) and the orientation rule (right) at T5. **h**, As in **g**, for training with a dense stimulus space ( $n = 3$  mice) to determine categorization behaviour closer to the category boundary. **i–k**, Psychometric curves for the three individual mice. The fraction chosen (fraction of Go choices) is shown along the relevant dimension (left) and the irrelevant dimension (right). Grey lines denote data from the T5 generalization session. Black lines denote data from the session with the dense stimulus space. **l**, As in **i**, showing the mean ( $\pm$  s.e.m.) across the three mice (shown in **i**, **j** and **k**) that were tested using the dense stimulus space. [Source data](#)

## Extended Data Fig. 4 Reconstruction of the location of imaging regions.

**a**, Top down view onto the craniotomy of M07 with prism implant (white square denotes the prism outline). A, anterior; P, posterior; Le, left; Ri, right. Scale bar, 0.5 mm. **b**, View through the prism with the position of an imaging field (white box), D, dorsal; V, ventral. Scale bar, 0.3 mm **c**, The imaging field in **b**, visualized with a two-photon microscope (red: structural marker mRuby2; green: functional marker GCaMP6m; image is the average of all frames of session T1). Scale bar, 30  $\mu$ m. **d**, Cropped images showing 12 example neurons across all imaging time points (T1–T8). **e**, The top triangle shows the correlation between cropped images of any two time points (average across all neurons). The bottom triangle shows the correlation after shuffling cell identities (control). **f**, Example sagittal brain section showing the position of the prism implant along the anterior-posterior axis. Scale bar, 1 mm. **g**, Schematic of cortical midline regions near the prism implant (ML 0.12), modified from Franklin & Paxinos<sup>64</sup>, figure 102, with permission from Academic Press (Copyright 2007). 3V, third ventricle; ACC, anterior cingulate cortex; CC, corpus callosum; D3V, dorsal third ventricle; MO, medial orbital cortex; OB, olfactory bulb; PL, prelimbic cortex; RSC, retrosplenial cortex; SC, superior colliculus. The centres of all imaging regions included in Fig. 2g,h are indicated by black dots. **h**, Fraction of category-selective cells for each imaged field of view (included in Fig. 2g,h). The black, hollow circles are imaging regions without category-selective cells. There was no clear relationship between the location of the imaging regions within mPFC and the fraction of category-selective cells. [Source data](#)

## Extended Data Fig. 5 Examples of single cells that became category-selective over the course of learning.

**a**, Example Go category-selective cell (from the mouse shown in Fig. 2). Top left, position of the cell in the two-photon image. Scale bar, 30  $\mu$ m. Bottom left, HLS map of the example region (hue: category identity of the presented stimuli; cyan, category 1; pink, category 2; lightness: response amplitude; saturation: selectivity for the stimulus category). Top middle,

before learning (T1),  $\Delta F/F$  traces aligned to stimulus onset. Scale bar, 100%  $\Delta F/F$ . Grey bar denotes stimulus presentation, 1.3 s. Bottom middle, mean inferred spike rate per stimulus. Right, after learning (T5). **b**, As in **a** for an example NoGo category-selective cell. **c**, Mean inferred spike rate per stimulus for six further category-selective cells from different mice. **d**, Response amplitude during the first presentation of each stimulus, averaged across all category-selective cells at T5. Left, green: Go-preferring neurons; right, red: NoGo-preferring neurons. **e**, Top row, 40 by 40 pixel cropped images showing a Go category-selective cell in the averaged two-photon imaged field of views (pseudo-coloured). Bottom row, mean inferred spike rate of the response of the example cell to the presented stimuli at each imaging time point. **f**, **g**, As in **e**, but for different, NoGo category-selective cells. [Source data](#)

### Extended Data Fig. 6 Category-tuning index distributions of all recorded field of views at T1, T5 and T8 and the overlap of populations of category-selective cells.

**a**, Category-tuning index before learning (T1, according to rule 1), after the mouse had learned rule 1 (T5, according to rule 1) and after it had learned to categorize stimuli according to rule 2 (T8, according to rule 2). Each imaging region is displayed individually. Individual cells are represented as dots. Only cells recorded at all imaging time points were included. Grey line indicates the threshold CTI value of 0.1 that was applied to classify cells as category-selective. **b**, Black, the fraction of overlap between category-selective groups found at T5 and T8 (Go stay/NoGo stay: Go/NoGo category-selective at both T5 and T8; Go switch: Go category-selective at T5 and NoGo category-selective at T8; NoGo switch: NoGo category-selective at T5 and Go category-selective at T8). Grey denotes 95% confidence intervals of chance population overlap (Methods). **c**, Go category-selective cells. Top row, Venn diagrams of the fraction of cells that were category-tuned only for rule 1 (R1), only for rule 2 (R2) or for both rules (area between R1 and R2). The highlighted part of the Venn diagram indicates which data are shown in the bottom row. Bottom row, mean stimulus response amplitude (inferred spike rate) after rule 1 (left) or rule 2 (right) was learned. **d**, As in **c**, for NoGo category-selective cells. [Source data](#)

## Extended Data Fig. 7 Individual neurons follow characteristic time courses of acquiring selectivity.

**a**, Left, scatter plot showing the difference in mean inferred spike rate between stimuli of the two categories, after learning the first rule (T5, *x* axis) and the rule-switch (T8, *y* axis) for individual Go category-selective cells at session T5 (blue) and T8 (orange). Right, histogram of the differences from unity of the distributions shown on the left.

$P_{T5} = 1.5 \times 10^{-8}$ ,  $P_{T8} = 3.5 \times 10^{-15}$ , two-tailed WMPSR ( $n_{T8} = 122$ ,  $n_{T5} = 156$ ). **b**, As in **a**, but showing the relative spike rate difference (normalized by the sum of inferred spike rate to category 1 and 2 stimuli) for individual Go category-selective neurons at T5 and T8.

$P_{T5} = 4.9 \times 10^{-21}$ ,  $P_{T8} = 1.5 \times 10^{-18}$ , two-tailed paired-samples *t*-test ( $n_{T5} = 156$ ,  $n_{T8} = 122$ ). **c**, As in **a**, but for NoGo category-selective cells at T5 and T8.  $P_{T5} = 9.1 \times 10^{-5}$ ,  $P_{T8} = 1.0 \times 10^{-12}$ , two-tailed WMPSR ( $n_{T5} = 57$ ,  $n_{T8} = 70$ ). **d**, As in **b**, but for NoGo category-selective cells.

$P_{T5} = 9.6 \times 10^{-8}$ ,  $P_{T8} = 1.7 \times 10^{-21}$ , two-tailed paired-samples *t*-test ( $n_{T5} = 57$ ,  $n_{T8} = 70$ ). **e**, Development of the spike rate difference up to T5, for individual Go category-selective neurons at T5. Before learning, baseline: T1. After learning the initial stimuli: T2. After learning categorization: T5. Grey lines denote individual neurons. Black line denotes the mean across cells. **f**, As in **e**, but for NoGo category-selective neurons.

**g**, Schematic showing predicted time courses for the acquisition of reward/choice (RC) selectivity, and category selectivity according to each rule (R1, R2). These predictors were fit to the time courses of individual neurons using linear regression in **h–k**. **h**, Left, mean ( $\pm$  s.e.m.) predictor weight of T5 Go category-selective neurons.  $P_{RC} = 3.2 \times 10^{-9}$ ,

$P_{R1} = 2.0 \times 10^{-7}$ ,  $P_{R2} = 0.12$ , two-tailed WMPSR tests, Bonferroni-corrected for three comparisons ( $n = 156$ ). Right, the predictor weights of individual neurons. Selectivity of Go-preferring neurons was best predicted by reward/choice, and also showed a category component. **i**, As in **h**, for T5 NoGo category-selective cells.  $P_{RC} = 0.03$ ,  $P_{R1} = 0.001$ ,  $P_{R2} = 0.03$ , two-tailed WMPSR tests, Bonferroni-corrected for three comparisons ( $n = 57$ ). Selectivity of NoGo-preferring neurons corresponded best to the time

course of acquiring category rule 1. **j**, As in **h**, for Go category-selective cells defined at T8  $P_{RC} = 0.03$ ,  $P_{R1} = 0.003$ ,  $P_{R2} = 7.8 \times 10^{-15}$ , two-tailed WMPSR tests, Bonferroni-corrected for three comparisons ( $n = 122$ ). **k**, As in **h**, for NoGo category-selective cells defined at T8  $P_{RC} = 0.09$ ,  $P_{R1} = 0.52$ ,  $P_{R2} = 8.0 \times 10^{-7}$ , two-tailed WMPSR tests, Bonferroni-corrected for three comparisons ( $n = 70$ ). The best predictor for both Go and NoGo preferring category-selective neurons after the rule-switch was the gradual acquisition of category rule 2. [Source data](#)

### Extended Data Fig. 8 Relation between motor behaviour and neuronal responses of category-selective cells.

**a**, Line histograms showing the count probability of behavioural (left) and neural (right) reaction times of individual mice. Behavioural reaction time (bRT) was measured as the time of the first lick after stimulus onset, neural reaction time (nRT) as the time of the neuronal response onset after stimulus onset. **b**, Left, scatter plot of bRT and nRT for every trial of every mouse in session T5.  $P = 2.3 \times 10^{-13}$ , rho = 0.08, Spearman's correlation ( $n = 9,348$  measured reaction times). Right, grey circles: scatter plot showing the average nRT (that is, the nRT averaged across all Go category-selective neurons, but separated per mouse and trial) versus the bRT per mouse and trial.  $P = 6.2 \times 10^{-6}$ , Pearson's  $r = 0.13$  ( $n = 1,156$  trials). The density of grey circles is indicated by the colour intensity (alpha value). Coloured circles: the overall mean nRT and bRT of each mouse.  $P = 0.51$ , Pearson's  $r = 0.26$  ( $n = 9$  mice). Dashed line denotes the unity line. **c**, CTI of Go (left) and NoGo (right) category-selective neurons, calculated for every imaging frame individually. Data show the period from 1 s before stimulus onset to 3 s after stimulus offset. Grey dashed line denotes the average time of first lick. Black line denotes the average period of stimulus presentation. **d**, Mean lick frequency in session T5, grouped by trial outcome (hits, misses, correct rejections and false alarms). Insets show the same data with inflated y axis. Black line denotes the average period of stimulus presentation. **e**, As in **d**, but showing the average running speed. **f**, Inferred spike rate of Go (left) and NoGo (right) category-selective neurons aligned to the onset of lick-bouts. Top row, lick-bouts detected within a trial. Bottom row, lick-bouts detected in the inter-trial-interval. Data are

mean  $\pm$  s.e.m. **g**, Inferred spike rate of Go category-selective neurons in session T5, grouped by trial outcome (hits, misses, correct rejections and false alarms). Black line denotes stimulus presentation. Data are mean  $\pm$  s.e.m. **h**, As in **g**, for NoGo category-selective neurons. **i**, Scatter plot showing the mean inferred spike rate in correct trials versus incorrect trials, for individual Go (green) and NoGo (red) category-selective neurons.  $P_{\text{Go}} = 1.0 \times 10^{-26}$ , Pearson's  $r_{\text{Go}} = 0.72$ ,  $P_{\text{NoGo}} = 4.1 \times 10^{-5}$ , Pearson's  $r_{\text{NoGo}} = 0.52$  ( $n_{\text{No}} = 156$ ,  $n_{\text{NoGo}} = 57$ ). Black line denotes the unity line. Line histogram shows the distribution of difference from unity separately for Go and NoGo-preferring neurons.  $P_{\text{Go}} = 6.6 \times 10^{-22}$ ,  $P_{\text{NoGo}} = 1.9 \times 10^{-5}$ , two-tailed WMPSR ( $n_{\text{Go}} = 156$ ,  $n_{\text{NoGo}} = 57$ ). [Source data](#)

### Extended Data Fig. 9 mPFC contains neural correlates of multiple task components.

**a**, Linear regression model, fitting the trial-averaged inferred spike rates of individual neurons at T5 (' $w_i$ ' denotes the predictor weight; category predictor 0: category 1; 1: category 2; [Methods](#)). **b**, Distribution of absolute choice predictor weight of all observed neurons, divided into low, middle and high weight groups with equal numbers of cells. **c**, Box plots of CTI distributions for the choice weight groups in **b**. Boxes show the first to third quartile of the distributions, and black line denotes the median. There was no significant difference between the distributions, showing that category selectivity is not observed exclusively in highly choice-correlated cells.  $P = 0.92$ , Kruskal–Wallis test comparing all groups, chi-squared = 0.158, d.f. = 2. **d**, Relative weights of linear regression predictors (category identity, choice, reward and running speed) of Go and NoGo category-selective cells at T5. Left, category, choice and reward predictors show a significant deviation from 0. Right, only the category predictor shows a significant difference from 0.  $P_{\text{Go-w1}} = 6.8 \times 10^{-5}$ ,  $P_{\text{Go-w2}} = 2.3 \times 10^{-7}$ ,  $P_{\text{Go-w3}} = 2.0 \times 10^{-14}$ ,  $P_{\text{Go-w4}} = 0.17$ ,  $P_{\text{NoGo-w1}} = 3.9 \times 10^{-5}$ ,  $P_{\text{NoGo-w2}} = 0.03$ ,  $P_{\text{NoGo-w3}} = 0.68$ ,  $P_{\text{NoGo-w4}} = 0.11$ , two-tailed WMPSR tests, Bonferroni corrected for four comparisons ( $n_{\text{Go}} = 156$ ,  $n_{\text{NoGo}} = 57$  cells). Grey boxes span the first to third quartile, black lines show the median. **e**, Distribution

of  $R^2$  values, black line at 0.05 denotes the cut-off for cells included in hierarchical clustering (resulting in 536 out of 2,306 neurons, largely excluding unresponsive neurons). **f**, Correlation of the  $R^2$  value of individual cells and their maximum average response to correct or incorrect trials of either category.  $P = 4.8 \times 10^{-121}$ , rho = 0.46, Spearman's correlation ( $n = 2,306$  cells). Grey line denotes the  $R^2$  cut-off shown in **e**, which eliminated mostly unresponsive neurons. **g**, Gap statistic of hierarchical clustering for varying cluster numbers. Arrow denotes the optimal number of clusters (nine clusters; [Methods](#)). Error bars denote the standard error of the gap statistic value. **h**, Principal component analysis of model weights shows cluster separation along the major axes of variance. Line histograms show distributions per cluster along PC1 and PC2 separately. Individual neurons (dots) are colour-coded by cluster identity. **i**, Top, dendrogram showing cluster linkage. Second row, for each neuron, relative weights of model predictors in each of the nine clusters. Third row, for each neuron, normalized responses in the four different trial outcomes. Fourth row, per cluster, mean normalized response to every stimulus.

[Source data](#)

### [Extended Data Fig. 10 Category selectivity throughout the task change and contributions of task-relevant parameters and uninstructed movements to explained response variance.](#)

**a**, Scatter plot showing the CTI of Go-preferring neurons having a  $CTI > 0.1$  in session T5 (blue) or session L/R (orange). Grey lines denote the CTI threshold used to determine category selectivity. **b**, As in **a**, for NoGo preferring cells. **c**, HLS maps of the example imaging region before (T5) and after (L/R) the task change (also shown in Fig. 4). Scale bar, 30  $\mu$ m. White circles indicate example cells in **d** and **e**. Hue: preferred category; lightness: response amplitude; saturation: category selectivity. **d**, Example Go-preferring neuron. Top, inferred spike rate for stimuli ordered along the relevant dimension (black), or the irrelevant dimension (blue). Inset, section from the HLS map in **c** showing the example cell. Bottom, average inferred spike rate per stimulus. Data are mean  $\pm$  s.e.m. **e**, As in **d**, but for a NoGo preferring example cell. **f**, Inferred spike rate of Go category-selective neurons (selected at T5), separated by stimulus/trial

outcome combination in the left/right choice task. Top row, category 1 (GoLeft is correct). Bottom row, category 2 (GoRight is correct). Grey, missed trials, no reward. Green, rewarded trials. Red, unrewarded trials. Black line, stimulus presentation. In each panel, ‘*n*’ indicates the total number of included trials (from nine mice). Data are mean ± s.e.m. **g**, As in **f**, for NoGo category-selective neurons (determined at T5). **h**, Category-selective neuronal responses, in absence of behavioural responses (missed trials in the left/right choice task). Inferred spike rate for each category presented in the left/right choice task, of Go category-selective neurons selected at T5.  $P = 3.3 \times 10^{-8}$ , two-tailed WMPSR ( $n = 407$ ). Data are mean ± s.e.m. **i**, As in **h**, but for NoGo category-selective neurons (determined at T5).  $P = 0.002$ , two-tailed WMPSR ( $n = 48$ ). **j**, Left and middle, schematic of the linear regression model, fitted to all trials of sessions T5 and 2AC combined. The average trial spike rate of each neuron was predicted by a weighted sum of the predictors: Category, Go, Reward, GoRight and GoLeft. Response vector and design matrix of example session. Right, significant normalized weights of all category-selective cells.  $P_{\text{Go-w1}} = 4.6 \times 10^{-19}$ ,  $P_{\text{Go-w2}} = 6.1 \times 10^{-37}$ ,  $P_{\text{Go-w3}} = 5.4 \times 10^{-43}$ ,  $P_{\text{Go-w4}} = 1.1 \times 10^{-23}$ ,  $P_{\text{Go-w5}} = 0.58$ ,  $P_{\text{NoGo-w1}} = 2.2 \times 10^{-6}$ ,  $P_{\text{NoGo-w2}} = 0.003$ ,  $P_{\text{NoGo-w3}} = 0.002$ ,  $P_{\text{NoGo-w4}} = 0.40$ ,  $P_{\text{NoGo-w5}} = 0.006$ , two-tailed WMPSR tests ( $n_{\text{Go}} = 407$ ,  $n_{\text{NoGo}} = 48$ ). **k**, Left, example cropped image of the body-imaging camera and the eye-imaging cameras, with overlaid marker positions (tracked using DeepLabCut<sup>41,42</sup>). Middle and right, schematics defining body and eye parameters derived from the tracked markers. **l**, Schematic showing predictors and the linear regression model used to fit the cells’ mean inferred spike rates per trial. **m**, Top, box plots showing the maximum predictive power ( $\text{cv}R^2$ ) of each model predictor ( $n = 9$  mice). Boxes show the first to third quartile; black line denotes the median. Bottom, box plots showing the unique contribution ( $\Delta R^2$ ) of each model predictor. **n**, Maximum predictive power ( $\text{cv}R^2$ , top) and the unique contribution ( $\Delta R^2$ , bottom) across Go category-selective neurons ( $n = 407$  neurons). Data are mean ± s.e.m. **o**, As in **n**, but for NoGo category-selective neurons ( $n = 48$  neurons). [Source data](#)

## Supplementary information

## Reporting Summary

### Video 1

Tracking of eye- and body features in the categorization task Recordings of both eyes and body of an example animal performing in session T5, shown is an excerpt of four trials. Top: Animal body, overlaid with DeepLabCut<sup>41,42</sup> annotated features. The red or green square appears during presentation of category 1 or 2, respectively. The blue square appears when the animal makes licks. Bottom left: Right eye, annotated with key features. Bottom right: Left eye, annotated with key features.

## **Source data**

[Source Data Fig. 1](#)

[Source Data Fig. 2](#)

[Source Data Fig. 3](#)

[Source Data Fig. 4](#)

[Source Data Extended Data Fig. 1](#)

[Source Data Extended Data Fig. 2](#)

[Source Data Extended Data Fig. 3](#)

[Source Data Extended Data Fig. 4](#)

[Source Data Extended Data Fig. 5](#)

[Source Data Extended Data Fig. 6](#)

[\*\*Source Data Extended Data Fig. 7\*\*](#)

[\*\*Source Data Extended Data Fig. 8\*\*](#)

[\*\*Source Data Extended Data Fig. 9\*\*](#)

[\*\*Source Data Extended Data Fig. 10\*\*](#)

## Rights and permissions

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

[\*\*Reprints and Permissions\*\*](#)

## About this article



Check for  
updates

## Cite this article

Reinert, S., Hübener, M., Bonhoeffer, T. *et al.* Mouse prefrontal cortex represents learned rules for categorization. *Nature* **593**, 411–417 (2021).

<https://doi.org/10.1038/s41586-021-03452-z>

[Download citation](#)

- Received: 25 November 2019
- Accepted: 12 March 2021
- Published: 21 April 2021
- Issue Date: 20 May 2021
- DOI: <https://doi.org/10.1038/s41586-021-03452-z>

## Comments

By submitting a comment you agree to abide by our [Terms](#) and [Community Guidelines](#). If you find something abusive or that does not comply with our terms or guidelines please flag it as inappropriate.

[Download PDF](#)

Advertisement

---

This article was downloaded by **calibre** from <https://www.nature.com/articles/s41586-021-03452-z>

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

Clofazimine broadly inhibits coronaviruses including SARS-CoV-2

[Download PDF](#)

- Article
- [Published: 16 March 2021](#)

# Clofazimine broadly inhibits coronaviruses including SARS-CoV-2

- [Shuofeng Yuan<sup>1,2,3</sup>](#)<sup>nal</sup>,
- [Xin Yin](#) [ORCID: orcid.org/0000-0003-2357-6718<sup>4,5</sup>](#)<sup>nal</sup>,
- [Xiangzhi Meng<sup>2</sup>](#)<sup>nal</sup>,
- [Jasper Fuk-Woo Chan](#) [ORCID: orcid.org/0000-0001-6336-6657<sup>1,2,3,6</sup>](#)<sup>nal</sup>,
- [Zi-Wei Ye<sup>2</sup>](#)<sup>nal</sup>,
- [Laura Riva<sup>5</sup>](#)<sup>nAff18</sup>,
- [Lars Pache](#) [ORCID: orcid.org/0000-0002-4193-1042<sup>5</sup>](#),
- [Chris Chun-Yiu Chan<sup>2</sup>](#),
- [Pok-Man Lai](#) [ORCID: orcid.org/0000-0002-7967-8126<sup>2</sup>](#),
- [Chris Chung-Sing Chan<sup>2</sup>](#),
- [Vincent Kwok-Man Poon<sup>2</sup>](#),
- [Andrew Chak-Yiu Lee](#) [ORCID: orcid.org/0000-0002-8432-3282<sup>2</sup>](#),
- [Naoko Matsunaga<sup>5</sup>](#),
- [Yuan Pu](#) [ORCID: orcid.org/0000-0002-7049-8452<sup>5</sup>](#),
- [Chun-Kit Yuen<sup>2</sup>](#),
- [Jianli Cao<sup>2</sup>](#),
- [Ronghui Liang<sup>2</sup>](#),
- [Kaiming Tang<sup>2</sup>](#),
- [Li Sheng](#) [ORCID: orcid.org/0000-0002-3139-8045<sup>7,8</sup>](#),
- [Yushen Du](#) [ORCID: orcid.org/0000-0002-8015-4207<sup>8</sup>](#),
- [Wan Xu<sup>7</sup>](#),
- [Chit-Ying Lau<sup>7</sup>](#),
- [Ko-Yung Sit](#) [ORCID: orcid.org/0000-0001-6289-663X<sup>9</sup>](#),
- [Wing-Kuk Au](#) [ORCID: orcid.org/0000-0002-6196-7317<sup>9</sup>](#),
- [Runming Wang](#) [ORCID: orcid.org/0000-0001-8525-5659<sup>10</sup>](#),
- [Yu-Yuan Zhang<sup>4</sup>](#),
- [Yan-Dong Tang<sup>4</sup>](#),

- [Thomas Mandel Clausen](#)<sup>11,12</sup>,
- [Jessica Pihl](#)<sup>11,13</sup>,
- [Juntaek Oh](#)<sup>11,14</sup>,
- [Kong-Hung Sze](#)<sup>1,2</sup>,
- [Anna Jinxia Zhang](#) [ORCID: orcid.org/0000-0002-5087-3614](#)<sup>1,2</sup>,
- [Hin Chu](#)<sup>1,2</sup>,
- [Kin-Hang Kok](#) [ORCID: orcid.org/0000-0003-3426-332X](#)<sup>1,2</sup>,
- [Dong Wang](#) [ORCID: orcid.org/0000-0002-2829-1546](#)<sup>11,14</sup>,
- [Xue-Hui Cai](#)<sup>4</sup>,
- [Jeffrey D. Esko](#)<sup>11,15</sup>,
- [Ivan Fan-Ngai Hung](#)<sup>16</sup>,
- [Ronald Adolphus Li](#)<sup>17</sup>,
- [Honglin Chen](#) [ORCID: orcid.org/0000-0001-5108-8338](#)<sup>1,2</sup>,
- [Hongzhe Sun](#) [ORCID: orcid.org/0000-0001-6697-6899](#)<sup>10</sup>,
- [Dong-Yan Jin](#) [ORCID: orcid.org/0000-0002-2778-3530](#)<sup>7</sup>,
- [Ren Sun](#) [7,8](#),
- [Sumit K. Chanda](#) [ORCID: orcid.org/0000-0001-9399-7927](#)<sup>5</sup> &
- [Kwok-Yung Yuen](#) [ORCID: orcid.org/0000-0002-2083-1552](#)<sup>1,2,3,6</sup>

[Nature](#) volume **593**, pages 418–423 (2021) [Cite this article](#)

- 53k Accesses
- 1 Citations
- 484 Altmetric
- [Metrics details](#)

## Subjects

- [Antivirals](#)
- [SARS-CoV-2](#)
- [Viral infection](#)

## Abstract

The COVID-19 pandemic is the third outbreak this century of a zoonotic disease caused by a coronavirus, following the emergence of severe acute respiratory syndrome (SARS) in 2003<sup>1</sup> and Middle East respiratory syndrome (MERS) in 2012<sup>2</sup>. Treatment options for coronaviruses are limited. Here we show that clofazimine—an

anti-leprosy drug with a favourable safety profile<sup>3</sup>—possesses inhibitory activity against several coronaviruses, and can antagonize the replication of SARS-CoV-2 and MERS-CoV in a range of in vitro systems. We found that this molecule, which has been approved by the US Food and Drug Administration, inhibits cell fusion mediated by the viral spike glycoprotein, as well as activity of the viral helicase. Prophylactic or therapeutic administration of clofazimine in a hamster model of SARS-CoV-2 pathogenesis led to reduced viral loads in the lung and viral shedding in faeces, and also alleviated the inflammation associated with viral infection. Combinations of clofazimine and remdesivir exhibited antiviral synergy in vitro and in vivo, and restricted viral shedding from the upper respiratory tract. Clofazimine, which is orally bioavailable and comparatively cheap to manufacture, is an attractive clinical candidate for the treatment of outpatients and—when combined with remdesivir—in therapy for hospitalized patients with COVID-19, particularly in contexts in which costs are an important factor or specialized medical facilities are limited. Our data provide evidence that clofazimine may have a role in the control of the current pandemic of COVID-19 and—possibly more importantly—in dealing with coronavirus diseases that may emerge in the future.

[Download PDF](#)

## Main

COVID-19 in humans has a broad clinical spectrum that ranges from mild to severe manifestations, with a mortality rate of about 2% worldwide<sup>4</sup>. The high transmissibility of SARS-CoV-2 has been attributed to a substantial proportion of mild or asymptomatic infections<sup>5,6</sup>. The genetically diverse coronavirus family—which is currently thought to comprise four genera (*Alphacoronavirus*, *Betacoronavirus*, *Gammacoronavirus* and *Deltacoronavirus*), infects birds and a variety of mammals (including bats)<sup>7</sup>. Within the past two decades, there have been three major outbreaks of disease caused by coronaviruses in humans. SARS-CoV was first detected in Guangdong (China) in 2002 and—with the aid of commercial air travel—spread rapidly and globally, and caused more than 8,000 cases with a 10% mortality rate<sup>1</sup>. In 2012, MERS-CoV may have evolved and spread from bats to humans through an intermediate host (camels); it has caused over 1,700 cases with a mortality rate of almost 40%, and—similar to SARS-CoV—air travel has enabled the global spread of this virus to 27 countries<sup>2,8</sup>.

There are no widely available specific antiviral therapies for coronaviruses in humans<sup>9</sup>. Remdesivir has been shown to exhibit pan-coronavirus inhibitory potential<sup>10</sup>, and has been granted approval by the US Food and Drug Administration (FDA) for the treatment of COVID-19 on the basis of the reduced time taken to

recover following treatment with this drug<sup>11</sup>. However, the therapy is suboptimal (particularly for patients with severe COVID-19), and it can only be administered intravenously to patients in hospital<sup>12,13</sup>. Thus, the development of additional therapeutic options is urgent, as is the establishment of combinatorial regimens to improve efficacy and reduce the potential for the emergence of drug-resistant variants. For example, the combination of IFN $\beta$ -1b, lopinavir–ritonavir and ribavirin has been shown to have beneficial antiviral effects in a randomized clinical trial<sup>14</sup>.

In efforts to accelerate the development of therapies for COVID-19, a library of known drugs that encompassed approximately 12,000 clinical-stage or FDA-approved small molecules was previously profiled<sup>15</sup>. This study focused on the antiviral mechanisms of action and in vivo efficacy of clofazimine, which is an FDA-approved molecule that was discovered as an anti-tuberculosis drug and was later used for the treatment of leprosy<sup>16</sup>. The effective concentration of clofazimine against SARS-CoV-2 in vitro (half-maximal effective concentration (EC<sub>50</sub>) of 0.31  $\mu$ M in Vero E6 cells) is clinically achievable with a single dose of 200 mg d<sup>-1</sup> (peak serum concentration ( $C_{max}$ ) of 0.86  $\mu$ M)<sup>17</sup>. A sufficiently high concentration of clofazimine, as multiples of its 90% effective concentration (EC<sub>90</sub>) in different cell lines (EC<sub>90</sub> of 0.81 to 2.35  $\mu$ M), is known to accumulate in plasma and lung tissue after several doses<sup>18</sup>. Here we report the protective effect of clofazimine against infection with SARS-CoV-2 or MERS-CoV in primary human cells and in a hamster model. In addition, clofazimine is an affordable treatment for patients with COVID-19, which may help to relieve the acute healthcare burden created by the pandemic in lower-income countries<sup>19</sup>.

## Clofazimine is a pan-coronavirus inhibitor

Clofazimine reduced MERS-CoV replication in Vero E6 cells, with an EC<sub>50</sub> of  $1.48 \pm 0.17 \mu$ M (Extended Data Fig. 1a). Immunofluorescence staining for MERS-CoV nucleocapsid (N) protein, combined with flow cytometry analysis, revealed the effective suppression of virus infection in human hepatocellular Huh7 cells upon clofazimine treatment (Extended Data Fig. 1b). We then characterized the antiviral activity of clofazimine in two cell lines that are more physiologically relevant to coronavirus diseases: cardiomyocytes derived from human embryonic stem cells and human primary small-airway epithelial cells, which robustly support SARS-CoV-2<sup>20</sup> and MERS-CoV<sup>21</sup> replication, respectively. Clofazimine treatment reduced the SARS-CoV-2 titre up to more than 3 log<sub>10</sub>, and the MERS-CoV titre by 2 log<sub>10</sub> (Fig. 1a).

Next, we assessed the antiviral activity of clofazimine in an ex vivo lung culture system and found that clofazimine potently antagonized viral replication in human lung tissues that reflect the primary site of SARS-CoV-2 and MERS-CoV replication

(Fig. 1b). To explore whether clofazimine confers cross-protection against other epidemic and seasonal coronaviruses, we performed viral-load reduction assays for SARS-CoV, hCoV-229E and hCoV-OC43 in corresponding cell lines that support virus replication. Viral yields in cell culture supernatants were decreased by about  $2 \log_{10}$  in Vero E6 cells infected with SARS-CoV, by about  $4 \log_{10}$  in human embryonic lung fibroblasts infected with hCoV-229E and by around  $3 \log_{10}$  in monkey BSC1 cells infected with hCoV-OC43 (Fig. 1c). Clofazimine showed negligible cytotoxicity in the matching cell lines as described above for pan-coronavirus inhibitory evaluation (Extended Data Fig. 1c). Overall, clofazimine exhibited broad-spectrum anti-coronavirus efficacy, and antagonized both SARS-CoV-2 and MERS-CoV replication in human primary cell and ex vivo lung models.

**Fig. 1: Clofazimine inhibits the replication of a broad spectrum of human-pathogenic coronaviruses in human cellular models.**

 **figure1**

**a**, Clofazimine inhibited SARS-CoV-2 (multiplicity of infection (MOI) of 0.1) replication in human primary cardiomyocytes (left) and MERS-CoV (MOI of 1.0) replication in human primary small-airway epithelial cells (right). Cell lysates were collected for viral load determination; viral genome copies were determined relative to *ACTB*. Data are mean  $\pm$  s.d.,  $n = 3$  biological replicates. Two-tailed Student's *t*-test. **b**,

Ex vivo human lung tissues were infected with SARS-CoV-2 or MERS-CoV-2 followed by treatment with clofazimine (10 µM), remdesivir (10 µM) or DMSO (0.1%). Supernatants were collected for the quantification of viral titre by plaque assay. Data are mean ± s.d.,  $n = 5$  biological replicates. Two-tailed Student's *t*-test. PFU, plaque-forming unit. **c**, Antiviral activity of clofazimine against SARS-CoV (MOI of 0.01, 48 h after infection) (left), HCoV-229E (MOI of 0.001, 72 h after infection) (middle) and HCoV-OC43 (MOI of 0.001, 72 h after infection) (right) in Vero E6, human embryonic lung fibroblasts and BSC1 cells, respectively. Viral load in the cell culture supernatant was quantified by qPCR with reverse transcription (RT–qPCR). Data are mean ± s.d.,  $n = 3$  biological replicates. One-way ANOVA followed by Sidak's post-test. All experiments were repeated twice for confirmation. All statistical analyses are compared with the DMSO-treated group (0 µM).

\*\*\*\* $P < 0.0001$ , \*\*\* $P < 0.001$ , \*\* $P < 0.01$ , \* $P < 0.05$ .

[Full size image](#)

## Effects on SARS-CoV-2 life cycle

We first evaluated the antiviral activity of clofazimine using a time-of-drug addition assay in a single infectious cycle. Treatment with clofazimine during inoculation strongly inhibited SARS-CoV-2 infection, which indicates that clofazimine inhibits viral entry. Clofazimine also blocked SARS-CoV-2 infection at a post-entry step, as evidenced by an observed reduction in viral replication when clofazimine was added at 5 h after infection (Extended Data Fig. 2a). To further evaluate the effect of clofazimine on viral entry, we used vesicular stomatitis virus (VSV)-based spike glycoprotein (S) pseudotyped virions. Clofazimine treatment reduced the infectivity of both SARS-CoV S and SARS-CoV-2 S pseudotyped virions in Vero E6 cells, but did not affect MERS-CoV S pseudotyped virus particles (Fig. 2a). We then examined a series of events to determine the precise step of SARS-CoV-2 entry that is blocked by clofazimine, and excluded possible effects of clofazimine on: (1) the cell-surface expression of the SARS-CoV-2 receptor angiotensin-converting enzyme 2 (ACE2) or the MERS-CoV receptor dipeptidyl peptidase 4 (DPP4) (Extended Data Fig. 2b) and (2) the disruption of binding between ACE2, or another attachment factor (heparan sulfate)<sup>22</sup>, and SARS-CoV-2 S (Extended Data Fig. 2c). Using a cell membrane fusion assay mediated by SARS-CoV-2 S<sup>23,24</sup>, we found that clofazimine inhibited cell fusion activity mediated by S. Specifically, we observed that fusion between effector cells that express S and target cells were effectively reduced after clofazimine addition, in a dose-dependent manner (Fig. 2c, Extended Data Fig. 2d). To confirm whether clofazimine also inhibits post-entry steps of viral replication, we evaluated the effect of clofazimine on the production of viral RNA by electroporating in vitro-transcribed viral RNA into Vero E6 cells, which bypasses clofazimine-mediated inhibition on the entry process and directly measures RNA synthesis (Extended Data Fig. 2e). As

expected, remdesivir reduced the synthesis of negative-stranded RNA in a dose-dependent manner (blue symbols in Fig. 2c). Viral RNA levels were also reduced by about  $1\text{--}1.5 \log_{10}$  in cells treated with clofazimine (red symbols in Fig. 2c). However, no substantial effect was observed on the translation of electroporated *GFP* mRNA (Extended Data Fig. 2f). We also found that clofazimine had no obvious effect on the activity of main protease and papain-like protease, which are responsible for the cleavage of viral polypeptide (half-maximal inhibitory concentration ( $\text{IC}_{50}$ ) of over 100  $\mu\text{M}$ ) (Extended Data Fig. 2g). Because a critical step in the replication and transcription of SARS-CoV-2 requires the formation of a stable complex of the SARS-CoV-2 helicase (nsp13) and holo-RNA-dependent RNA polymerase (RdRp)<sup>25</sup>, we investigated the effect of clofazimine on this step, and found that clofazimine inhibited the unwinding activity of nsp13 using either a double-stranded DNA or double-stranded RNA substrate (Fig. 2d). However, clofazimine did not block the primer elongation activity that is executed by holo-RdRp (Extended Data Fig. 2h). Collectively, these results demonstrated that clofazimine exerts antiviral activity by targeting several steps in SARS-CoV-2 replication, including interference with S-mediated cell fusion as well as viral helicase activity.

**Fig. 2: Clofazimine interferes with several steps of the virus life cycle.**

---

 **figure2**

**a**, VSV-based pseudotyped viral particle assay. Vero E6 cells that were pretreated with indicated compounds were infected with the SARS-CoV-2 S- (left), SARS-CoV S- (middle) or MERS-CoV S- (right) pseudotyped particles. Luciferase signals were quantified at 24 h after infection. Mean  $\pm$  s.e.m.,  $n = 6$  independent experiments. One-way ANOVA followed by Dunnett post-test. MDL28170, a known coronavirus entry inhibitor<sup>15</sup>, was used as a positive control. RLU, relative light unit. **b**, Clofazimine inhibits membrane fusion mediated by SARS-CoV-2 S. Vero cells cotransfected with SARS-CoV-2 S and EGFP plasmids were added to the nontransfected Vero cells to induce membrane fusion. Confocal images were acquired at 48 h after transfection. Scale bars, 400  $\mu$ m. Representative images selected from a pool of images captured in

two independent experiments. **c**, In vitro-transcribed viral RNA replication assay. Remdesivir or clofazimine were added at the indicated doses. Negative-stranded RNA was then quantified at 12 h after electroporation. Error bars are s.e.m.,  $n = 3$  independent experiments. Two-tailed Student's *t*-test. **d**, Titration of double-stranded (ds)DNA and dsRNA unwinding activity of the SARS-CoV-2 helicase (nsp13) by clofazimine using an assay based on fluorescence resonance energy transfer. The black curve represents a positive-control inhibitor (ranitidine bismuth citrate) using a DNA-based substrate. Data are mean  $\pm$  s.d.,  $n = 3$  biological replicates. The experiments were repeated twice for confirmation. All statistical analyses are compared with the DMSO-treated or non-treatment group. \*\*\* $P < 0.001$ , \*\* $P < 0.01$ , \* $P < 0.05$ ; NS, not significant.

[Full size image](#)

## Transcriptional analysis after treatment

We used RNA sequencing (RNA-seq) to profile transcriptome-wide changes in human Caco-2 cells infected with SARS-CoV-2, in which clofazimine exhibited antiviral potency comparable to that of remdesivir (Extended Data Fig. 3a). We performed transcriptional analysis on Caco-2 cells that were infected or uninfected and that were subsequently treated with clofazimine or dimethyl sulfoxide (DMSO; mock treatment) for 3 or 6 h (Extended Data Fig. 3b). At 3 h after infection, clofazimine treatment caused an overall transcriptome shift towards the mock-treatment control group (Extended Data Fig. 3c), which corroborates the hypothesis that the drug interferes with the early stage of the virus life cycle. At 6 h after infection, there were 607 and 448 genes upregulated and downregulated, respectively, by SARS-CoV-2 infection (Supplementary Tables 1, 2). The RNA level of more than 90% of these genes was reverted by clofazimine treatment, which indicates that clofazimine treatment abrogates the transcriptomic changes that are induced by SARS-CoV-2 infection. This is consistent with a principal components analysis of the dataset, which showed that treatment with clofazimine for 6 h after infection caused a marked shift towards the mock treatment (Extended Data Fig. 3c, d). Clofazimine treatment for 6 h in the absence of infection upregulated the expression of transcription factors that are critical for the immediate–early cellular response, including those of the AP-1, SMAD and MAFF families (Extended Data Fig. 3e). When clofazimine was applied to infected cells for 6 h, we observed an enrichment of upregulated genes that are associated with innate-immunity-related pathways, including MAPK, interleukin and TNF responses (Extended Data Figs. 3f, 4a, b). These results suggest that clofazimine may also rewire the transcriptional landscape to prime the innate-immunity-related pathways.

## Prophylactic and therapeutic activities *in vivo*

Previous pharmacokinetics studies have revealed that, in humans, co-administration of a 200 mg dose of clofazimine with food results in a  $C_{\max}$  of  $0.41 \text{ mg l}^{-1}$  (equivalent to  $0.86 \mu\text{M}$ ) at 8 h. Here we used a golden Syrian hamster model<sup>26</sup> to determine the in vivo antiviral efficacy of clofazimine. Because administration of clofazimine with a high-fat meal provides better bioavailability<sup>27</sup>, we delivered the drug through the oral route using corn oil as vehicle (Fig. 3a).

**Fig. 3: Prophylactic and therapeutic treatment with clofazimine reduces disease in a hamster model of SARS-CoV-2 infection.**

 figure3

**a**, The prophylactic treatment used oral administration of clofazimine, given at 3, 2 and 1 days before infection, followed by virus challenge at day 0; therapeutic administration of clofazimine was performed at 1, 2, and 3 days after infection. Tissue samples were collected at indicated days after infection. Remdesivir was included as a control in the therapeutic regimen. **b, c**, Daily body weights of the hamsters under prophylactic (**b**) or therapeutic (**c**) treatment. Data are mean  $\pm$  s.e.m.,  $n = 3$  hamsters per group. Two-way ANOVA followed by Sidak's post-test for prophylactic groups, and two-way ANOVA followed by Tukey's post-test for therapeutic groups.  $P$  value indicates clofazimine versus vehicle group. **d**, Viral yield in hamster lung tissue, after prophylactic ( $n = 5$  hamsters per group) or therapeutic treatment ( $n = 11$  hamsters (each remdesivir and clofazimine group),  $n = 13$  hamsters (vehicle group)), which were collected at 4 days after infection and titrated by plaque assays. **e**, Hamster nasal washes collected on 4 days after infection were subjected to live virus titration by plaque assays ( $n = 5$  hamsters per group). **f**, Hamster faeces were freshly collected at 4 days after infection and subjected to SARS-CoV-2 viral copy detection by RT-qPCR assays ( $n = 5$  hamsters per group). For purposes of statistical analysis, a value of about 10–100 was assigned for any data point below the detection limit (dotted line). **g**, The IL-6 level in hamster serum was quantified at 4 days after infection.  $n = 5$  hamsters for each prophylaxis group and  $n = 10$  hamsters for each therapeutic group. All data in **d–g** are mean  $\pm$  s.d. Two-tailed Student's *t*-test for prophylaxis groups and one-way

ANOVA followed by Dunnett's post-test for therapeutic groups, comparing with the vehicle group (black symbols). \*\*\*\* $P < 0.0001$ , \*\*\* $P < 0.001$ , \*\* $P < 0.01$ , \* $P < 0.05$ , NS, not significant.

[Source data](#)

[Full size image](#)

After clofazimine administration in a prophylactic regimen, we observed a reduction in the loss of body weight after infection with SARS-CoV-2. When clofazimine was given after virus challenge, the time taken for body weight to return to pre-infection levels was shortened (Fig. 3b, c). At 4 days after infection (when viral loads peaked and there were substantial histopathological changes), clofazimine decreased the virus plaque-forming units in lung tissues by about 1–2 log<sub>10</sub> (Fig. 3d). We consistently confirmed the suppression of viral genome copies of SARS-CoV-2 in the lungs in hamsters treated with clofazimine (Extended Data Fig. 5a). In addition, we evaluated the capacity of clofazimine to diminish SARS-CoV-2 shedding in nasal wash and faecal samples. At 4 days after infection and after remdesivir or clofazimine therapy, we did not detect any significant ( $P > 0.05$ , one-way analysis of variance (ANOVA)) reduction in infectious virus titre in the hamster nasal wash, whereas an average of a twofold reduction was achieved with prophylactic administration of clofazimine (Fig. 3e). Irrespective of whether clofazimine was administered prophylactically or therapeutically, we found substantially lower viral loads in faeces at 4 days after infection compared with the vehicle-treated group. By contrast, remdesivir did not alleviate virus shedding from gastrointestinal tract (Fig. 3f). Overall, we demonstrate in vivo in a hamster model that clofazimine antagonizes SARS-CoV-2 replication in the lung and reduces virus shedding in faeces.

To ascertain whether clofazimine improves disease prognosis, we determined serum levels of IL-6, which have previously been shown to correlate with respiratory failure and adverse clinical outcomes<sup>28</sup>. In general, we detected substantially decreased levels of serum IL-6 in both the clofazimine-treated and remdesivir-treated groups (Fig. 3g). It has previously been reported that clofazimine may inhibit lymphocyte function in cell culture<sup>29</sup>. To investigate this in the context of an immune response induced by SARS-CoV-2, we collected sera from hamsters at 14 days after infection and found that the levels of antibody responses triggered in vehicle-treated and clofazimine-treated groups were similarly high, which indicates that clofazimine does not induce detectable suppression of the humoral immune response of B lymphocytes (Extended Data Fig. 5b). We also performed RNA-seq of hamster lung tissues (Extended Data Fig. 6a). Out of the 73 upregulated genes after prophylactic treatment with clofazimine (Supplementary Table 3), 34 (46.6%) were mapped to biological processes related to the immune response (Extended Data Fig. 6b). Of these 34 genes, 13 were mapped to

the ‘leukocyte differentiation’ Gene Ontology category—including two major histocompatibility complex class-II molecules, H2-Aa and H2-Ab1 (which are labelled with an asterisk in Extended Data Fig. [6c](#)). In addition, transcription factors—including *Fos*, *Junb* and *Egr1*—were also upregulated, which is consistent with our transcriptomic analysis of Caco-2 cells treated with clofazimine (Extended Data Fig. [6c](#)). Notably, most of the immune-response-related genes were not changed after administering clofazimine to uninfected hamsters, which indicates that clofazimine priming of the host response is dependent on SARS-CoV-2 infection and is not likely to result in the undue activation of the host immune system (Extended Data Figs. [6d](#), [e](#)).

To determine the severity of lung damage, we performed histological examination of lung tissues that were stained with haematoxylin and eosin. The lungs of mock-treated control hamsters showed severe pathological changes, which were evidenced by large areas of consolidation as well as cell infiltrations in the endothelium of blood vessels, and peribronchiolar regions. By contrast, clofazimine- or remdesivir-treated lungs exhibited an improved morphology and milder infiltrations (Extended Data Fig. [7a, b](#)). Taken together, our results suggest that clofazimine conferred protection against SARS-CoV-2 challenge in the hamster model by reducing virus replication and associated inflammatory dysregulation in the host.

## Antiviral synergy with remdesivir

Remdesivir is considered the current standard of care for the treatment of COVID-19. We found that co-application of clofazimine and remdesivir affects SARS-CoV-2 replication in a manner that extends beyond the additive activity predicted by the Bliss independence model (maximal Bliss synergy score of 44.28), which indicates that these two drugs operate in a synergistic antiviral relationship (Extended Data Fig. [8a, b](#)). The addition of 1.25  $\mu$ M clofazimine (fourfold EC<sub>50</sub>) in an in vitro assay resulted in a nearly 20-fold decrease in the concentrations of remdesivir required to inhibit viral replication by 90% (Extended Data Fig. [8c](#)) and—importantly—the combination of drugs did not elicit additional cellular cytotoxicity (Extended Data Fig. [8d](#)).

To explore the antiviral synergy of these two drugs *in vivo* and to recapitulate the scenario in which most patients with COVID-19 will be treated after the onset of disease, we gave SARS-CoV-2-infected hamsters oral clofazimine and intraperitoneal remdesivir together: the first doses were given 24 h after virus challenge. Experimentally, we gave 1.5 mg per kg body weight remdesivir (tenfold lower than the standard dosing) as well as a reduced clofazimine dose (15 mg per kg bodyweight, corresponding to a 100 mg dose in humans) (Fig. [4a](#)). An improvement in weight loss was achieved in the combinatorially treated group at 3 days after infection, compared with the groups treated with vehicle control or low-dose remdesivir alone (Fig. [4b](#)).

Monotherapy using low-dose remdesivir caused a marginal reduction of virus titre in the lungs. However, the additional application of clofazimine not only exhibited potent synergy in terms of viral load (Fig. 4c), but also restricted virus replication in the focal bronchiolar epithelial cells from spreading to alveolar areas (Extended Data Fig. 9a). The antiviral synergy also suppressed virus shedding in the nasal wash (Fig. 4d), which was not achievable with therapeutic treatments using remdesivir or clofazimine individually (Fig. 3e). Immunofluorescence staining of hamster nasal turbinates showed that the combinatorial treatment effectively diminished N antigen expression in the epithelium, providing further support for the synergy between the drugs (Extended Data Fig. 9b,c). Taken together, the antiviral synergy between low-dose remdesivir and clofazimine effectively improved viral control, leading to reduced body-weight loss, a suppressed pulmonary virus titre and nasal virus shedding, as well as decreased drug dosages.

**Fig. 4: Clofazimine exhibits antiviral synergy with remdesivir in hamsters.**

 figure4

**a**, Experimental design of in vivo study. Oral clofazimine (15 mg per kg body weight) and/or intraperitoneal remdesivir were administrated at the indicated days after infection. Vehicle-treated control hamsters received oral administration of corn oil and an intraperitoneal injection of 2% DMSO in 12% sulfobutyl ether- $\beta$ -cyclodextrin (SBE- $\beta$ -CD). The standard dose of remdesivir is 15 mg per kg body weight, a low-dose remdesivir is 1.5 mg per kg; the low dose was used in the combined treatment. **b**, Daily body weights of the hamsters ( $n = 5$  hamsters per group). Data are mean  $\pm$  s.d. Two-way ANOVA followed by Tukey's test. **c**, Virus in lung tissue was titrated by plaque assay. One-way ANOVA followed by Dunnett's post-test ( $n = 5$  hamsters per group). **d**, Virus titre in nasal wash was determined by plaque assays. One-way ANOVA followed by Dunnett's post-test, compared with the combinatorial group ( $n = 5$  hamsters per group). \*\* $P < 0.01$ , \* $P < 0.05$ , NS, not significant.

[Source data](#)

[Full size image](#)

## Discussion

Clofazimine was first used to treat leprosy in 1969 and gained FDA approval in 1996<sup>30</sup>. It is an orally bioavailable drug that is included in the WHO (World Health Organization) Model List of Essential Medicines. It is generally well-tolerated, and is recommended as a WHO group-C drug for the treatment of multidrug-resistant and extensively drug-resistant tuberculosis<sup>31</sup>.

In SARS-CoV-2 infection, a delayed innate immune response may result in an uncontrolled cytokine storm<sup>32,33</sup>. The effect of clofazimine on rewiring the transcriptional landscape of the cell towards an antiviral status may be important in the COVID-19 setting, and understanding the contribution of this activity to in vivo disease amelioration can provide insights into its potential to improve viral control through the enhancement of innate immune activities. Elucidation of how clofazimine treatment may balance the regulation of innate and adaptive immune responses will be important to understanding its potential clinical efficacy. Further modification of clofazimine to improve its pharmacokinetics and randomized clinical trials to ascertain its clinical performance are warranted.

## Methods

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.

### Cells and viruses

The different cell lines and primary cells that were used in this study were chosen according to their high sensitivity to replication of a corresponding coronavirus. Human hepatoma Huh7 (JCRB, 0403) cells, human colon Caco-2 cells (ATCC, HTB-37), monkey Vero E6 cells (ATCC, CRL-1586) and monkey kidney BSC-1 cells (ATCC, CCL-26) were maintained in DMEM culture medium supplemented with 10% heat-inactivated FBS, 50 U ml<sup>-1</sup> penicillin and 50 µg ml<sup>-1</sup> streptomycin. Human embryonic lung fibroblasts were developed in-house. Human primary small-airway epithelial cells (ATCC, PCS-301-010) were cultured with airway epithelial cell basal medium, according to the manufacturer's protocol. Ventricular cardiomyocytes were differentiated from the human embryonic stem cell HES2 (ESI) maintained in mTeSR1 medium (STEMCELL Technologies)<sup>34</sup>. In brief, HES2 cells were dissociated with Accutase (Invitrogen) into single-cell suspensions on day 0. Cells were seeded on low-attachment culture vessels (Corning) and cultured in mTeSR1 medium supplemented with 40 µg ml<sup>-1</sup> Matrigel, 1 ng ml<sup>-1</sup> BMP4 (Invitrogen) and 10 µM Rho kinase inhibitor (ROCK) (R&D) under hypoxic environment with 5% O<sub>2</sub>. From day 1 to 3, cells were cultured in StemPro34 SFM (Invitrogen) with 50 µg ml<sup>-1</sup> ascorbic acid

(AA) (Sigma), 2 mM Gluta-MAX (Invitrogen), 10 ng ml<sup>-1</sup> BMP4, and 10 ng ml<sup>-1</sup> human recombinant activin-A (Invitrogen). From day 4 to day 7, 5 µM Wnt inhibitor IWR-1 (Tocris) was added. From day 8 to day 14, cells were cultured under normoxia in RPMI 1640 medium (Invitrogen) supplemented with 2 mM Gluta-MAX, 1× B-27 supplement (Invitrogen) and 50 µg ml<sup>-1</sup> AA. The cells were then dissociated with Accutase and seeded as a monolayer in desired culture vessels for 3 days before infections. The SARS-CoV-2 HKU-001a strain (GenBank accession number: MT230904) was isolated from the nasopharyngeal aspirate specimen of a patient who was laboratory-confirmed to have COVID-19 in Hong Kong<sup>26</sup>. The SARS-CoV-2 isolate USA-WA1/2020 was deposited by the Centers for Disease Control and Prevention, and obtained through BEI Resources. The MERS-CoV strain (HCoV-EMC/2012) was a gift from R. Fouchier. The archived clinical strains of SARS-CoV (HCoV-OC43 and HCoV-229E) were obtained from the Department of Microbiology of the University of Hong Kong (HKU)<sup>35</sup>. All experiments involving live SARS-CoV, SARS-CoV-2 and MERS-CoV followed the approved standard operating procedures of the biosafety level 3 facility at the University of Hong Kong and Sanford Burnham Prebys Medical Discovery Institute, as was previously described<sup>15</sup>.

## Antiviral evaluation in human ex vivo lung tissues

Human lung tissues for ex vivo studies were obtained from patients who were undergoing surgical operations at Queen Mary Hospital (Hong Kong), as previously described<sup>36</sup>. The donors and/or parents of the donor gave written consent, as approved by the Institutional Review Board of the HKU/Hospital Authority Hong Kong West Cluster (UW13-364). The freshly obtained lung tissues were processed into small rectangular pieces and were rinsed with advanced DMEM/F12 medium (Gibco) supplemented with 2 mM of HEPES (Gibco), 1× GlutaMAX (Gibco), 100 U ml<sup>-1</sup> penicillin and 100 µg ml<sup>-1</sup> streptomycin. The specimens were infected with SARS-CoV-2 HKU-001a or MERS-CoV with an inoculum of 1 × 10<sup>6</sup> PFU ml<sup>-1</sup> at 500 µl per well. After 2 h, the inoculum was removed, and the specimens were washed 3 times with PBS. The infected human lung tissues were then cultured in 1 ml of advanced DMEM/F12 medium with 2 mM HEPES (Gibco), 1× GlutaMAX (Gibco), 100 U ml<sup>-1</sup> penicillin, 100 µg ml<sup>-1</sup> streptomycin, 20 µg ml<sup>-1</sup> vancomycin, 20 µg ml<sup>-1</sup> ciprofloxacin, 50 µg ml<sup>-1</sup> amikacin and 50 µg ml<sup>-1</sup> nystatin. Supernatants were collected at 24 h after inoculation for plaque assays.

## Antiviral assessment in a hamster model of SARS-CoV-2 infection

Male and female Syrian hamsters, aged 6–10 weeks old, were obtained from the Chinese University of Hong Kong Laboratory Animal Service Centre through the HKU Centre for Comparative Medicine Research. The hamsters were kept in biosafety

level 2 housing and given access to standard pellet feed and water ad libitum, as was previously described<sup>26</sup>. All experimental protocols were approved by the Animal Ethics Committee in the HKU (CULATR) and were performed according to the standard operating procedures of the biosafety level 3 animal facilities (reference code: CULATR 5370-20). Experimentally, each hamster was intranasally inoculated with  $10^5$  PFU of SARS-CoV-2 in 100 µl PBS under intraperitoneal ketamine (200 mg per kg) and xylazine (10 mg per kg) anaesthesia.

To simulate the prescribed human dosage (200 mg per day), an equivalent hamster dose of 25 mg per kg per day was converted on the basis of body surface area. Specifically, 25 mg per kg (hamster)  $\times$  0.13 (conversion factor) = 3.25 mg per kg (human equivalent dose); and a 60-kg human requires 3.25 mg per kg  $\times$  60 kg = 195 mg clofazimine per day. Prophylactic treatment used oral administration of clofazimine given 3, 2 and 1 days before infection (25 mg per kg each time), followed by virus challenge at day 0; therapeutic post-exposure and oral administration of clofazimine (Sigma-Aldrich, C8895) was performed on 1, 2 and 3 days after infection (25 mg per kg each time) with the first dosage given at 24 h after inoculation. Clofazimine was delivered using corn oil (Sigma-Aldrich, C8267) as vehicle. Remdesivir was included as a positive-control drug and dosed at 15 mg per kg via the intraperitoneal route on the basis of its effective dosage in SARS-CoV-infected mice<sup>10</sup>. Remdesivir (15 mg per kg, MedChemExpress) was prepared as 100 mg ml<sup>-1</sup> stock in DMSO and further diluted using 12% SBE-β-CD before intraperitoneal injection. Hamsters receiving pure corn oil (oral) and 2% DMSO in 12% SBE-β-CD (intraperitoneal) were used as the vehicle-control group. Hamsters were killed at 4 days after infection for virological and histopathological analyses. Viral yield in the lung tissue homogenates and/or faeces was detected by plaque assay and/or RT-qPCR. Nasal washes were collected to examine virus shedding from the respiratory tract. Hamsters were treated with isoflurane lightly, and then 200 µl of PBS was injected into one nasal opening while turbid wash was collected from the other one without any blood contamination. The nasal wash was filtered through 0.22-µm filters before being subjected to plaque assay. An ELISA kit was used to determine the amount of IL-6 in the hamster sera at 4 days after infection, according to the manufacturer's recommendations (ELISAGenie, HMFI0001). Tissue pathology of infected hamsters was examined by haematoxylin and eosin (H&E) staining in accordance with the established protocol<sup>37</sup>. On day 14 after infection, an enzyme immunoassay was used to determine the antibody titre of hamster sera against SARS-CoV-2 N antigen. In brief, 96-well immune plates (Nunc) were coated with 100 µl per well (0.1 µg per well) of SARS-CoV-2 N in 0.05 M NaHCO<sub>3</sub> (pH 9.6) overnight at 4 °C. After blocking, 100 µl of heat-inactivated serum samples was serial-diluted before being added to the wells and incubated at 37 °C for 1 h. The attached antibodies were detected using horseradish-peroxidase-conjugated rabbit anti-hamster IgG antibody (Thermo Fisher Scientific; A18895; 1:2,000). The reaction was developed by adding diluted 3,3',5,5'-

tetramethylbenzidine single solution (TMB) (Invitrogen) and stopped with 0.3 N H<sub>2</sub>SO<sub>4</sub>. The optical density was read at 450/620 nm using a microplate reader.

## RNA-seq analysis

FASTQ files from RNA-seq were quality-examined by FastQC (v.0.11.7) (<https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>). Reads were processed by cutadapt to remove reads with low quality and to trim adapters. For RNA-seq on Caco-2 cells, trimmed reads were aligned to hg38 reference genome and NCBI SARS-CoV-2 reference genome (NC\_045512.2) using TopHat<sup>38</sup>(v.2.1.1); and RNA-seq data from hamster lung tissues were mapped to MesAur1.0 (GCA\_000349665.1), downloaded from Ensembl. Reads assigned to each gene were counted by featureCounts<sup>39</sup> (v.2.0.1) with human refseq gene sets as references for Caco-2 cells, and gene annotation of golden hamster from Ensembl database for the hamster dataset. Genes without at least 1 read mapped on average in each sample were considered undetectable and were filtered out. Read counts were normalized by the trimmed mean of *M*-values method and differential expression was calculated using R package edgeR (v.3.28.1) and genewise negative binomial generalized linear models with quasi-likelihood tests method was used for statistical tests. Cut-offs imposed for differential expression analysis were set as a false discovery rate (FDR) of 0.05 and fold change >2 or <0.5. The pathway analysis was performed by R package clusterProfiler<sup>40</sup> (v.3.14.3) and Metascape<sup>41</sup>. Heat maps were plotted using R package pheatmap (v.1.0.12) (<http://CRAN.R-project.org/package=pheatmap>). Other plots were generated by R package ggplot2 (v.3.3.0) (<https://ggplot2.tidyverse.org>). PCA analysis was performed by R package factoextra (1.0.7).

## Pseudotyping of VSV and pseudotype-based inhibition assay

VSV pseudotyped with the S of MERS-CoV, SARS-CoV or SARS-CoV-2 were generated as previously reported, with some modifications<sup>42</sup>. In brief, BHK-21/WI-2 cells (Kerafast) overexpressing the S were inoculated with VSV-G pseudotyped ΔG-luciferase VSV (Kerafast). After a 2-h inoculation at 37 °C, the inoculum was removed and cells were refed with DMEM supplemented with 5% FBS and VSV-G antibody (I1, mouse hybridoma supernatant from CRL-2700; ATCC; 1:100). Pseudotyped particles were collected at 24 h after inoculation, then centrifuged at 1,320g to remove cell debris and stored at -80 °C until use.

To determine the effect of the compounds on viral entry, Vero E6 cells were treated with clofazimine at a concentration of 2.5 μM for 1 h before inoculation with respective pseudotyped VSV. After a 2-h inoculation in the presence of the compounds, the inoculum was removed, and cells were refed with fresh medium for

further culture. The activity of firefly luciferase was measured using bright-Glo luciferase assay (Promega) for quantitative determination at 16 h after transduction.

## The effect of clofazimine on SARS-CoV-2 viral replication

The full-length SARS-CoV-2 viral RNA transcripts were in vitro-synthesized from an infectious clone of SARS-CoV-2 (provided by P.-Y. Shi) according to a recently published protocol<sup>43</sup>. Ten µg of total RNA transcripts and 5 µg SARS-CoV-2 N gene transcript were mixed with Vero E6 cells stably expressing SARS-CoV-2 N and then added into a 0.2-cm cuvette for nucleofection with the 4D-Nucleofector Core Unit (Lonza) using pulse code V-001. Immediately after electroporation, 1,000 µl of prewarmed medium was added to the cuvette and cells were subsequently aliquoted into 384-well plates. Two hours after seeding, compounds at different concentrations were added into each well. At 12 h after electroporation, intracellular and viral RNA was purified from the treated cells with TurboCapture 384 mRNA Kit (Qiagen) in accordance with the manufacturer's instructions. The purified RNA was subjected to first-strand cDNA synthesis using the high-capacity cDNA reverse transcription kit (Applied Biosystems) with the following primer (TagRdRp-F: 5'-CGGTCATGGTGGC GAATAACCCTGTGGG TTTTACACTTAA-3'). Real-time PCR analysis was performed using TaqPath 1-step RT-qPCR Master Mix (Applied Biosystems). The following primers and probe were used for negative-stranded RNA detection: Tag-F: 5'-CGGTCATGGTGGCGAATAACCCTGT-3', ORF1ab-R: 5'-ACGATTGTGCATCAGCTGA-3', ORF1ab-P: 5'-6FAM-CCGTCTGCGGT A TGTGGAAAGGTTATGG -BHQ1-3'). In parallel, 5 µg of the in vitro-transcribed GFP mRNA (StemMACS eGFP mRNA, 130-101-114) was electroporated into Vero E6 cells. At 2 h after seeding, cells were treated either with clofazimine or remdesivir, and then cultured for a further 24 h. The GFP signal was measured by flow cytometry analysis using FlowJo (v.10.0.7).

## Detection of S binding against ACE2 or heparin

Binding between the purified SARS-CoV-2 S and the cellular entry factors ACE2 or heparin were detected by ELISA, as was previously described<sup>22</sup>. High-binding microtitre plates were coated with heparin–BSA (100 ng per well) or recombinant ACE2 (200 ng per well) overnight at 4 °C. The plates

were then blocked for 3 h at 37 °C with TSM buffer (20 mM Tris buffer, pH 7.4, containing 150 mM NaCl, 2 mM MgCl<sub>2</sub>, 2 mM CaCl<sub>2</sub>, 0.05% Tween-20 and 1% BSA). Next, 10 nM biotinylated SARS-CoV-2 S in a dilution of clofazimine (10–0.1 µM) in TSM buffer was added to the plates in triplicate. Bound biotinylated protein was detected by adding Avidin-HRP (405103, BioLegend) diluted 1:2,000 in TSM buffer. Finally, the plates were developed with TMB turbo substrate for 5–15 min. The reaction was quenched using 1 M sulfuric acid and the absorbance was measured at 450 nm.

### S-mediated membrane fusion assay

SARS-CoV-2 S-mediated cell–cell fusion assay was performed as was previously established, with some modifications<sup>22</sup>. Vero cells were cotransfected with 1 µg SARS-CoV-2 S plasmid and 0.4 µg EGFP plasmid for 8 h. In another 24-well plate, Vero cells (that is, target cells expressing ACE2) were washed with PBS before being cocultured with the effector cells (that is, Vero cells cotransfected with S and EGFP), in the presence of the indicated concentration of clofazimine. After another 48 h, S-mediated membrane fusion, as reflected by the green fluorescence area, was evaluated using confocal imaging. Vero cells transfected with EGFP, without S transfection, were included as a negative control.

### Time-of-addition assay

A time-of-drug-addition assay was performed to investigate which stage of the SARS-CoV-2 life cycle it is that clofazimine interferes with, as previously described<sup>15</sup>. In brief, Vero E6 cells were seeded in 96-well plates ( $4 \times 10^4$  cells per well). The cells were infected by SARS-CoV-2 USA-WA1/2020 at an MOI of 1.5 and then incubated for additional 1 h. The viral inoculum was then removed, and the cells were washed twice with PBS. At 1 h after inoculation (that is, after entry), clofazimine at a concentration of 5 µM was added to the infected cells at time points indicated, followed by incubation at 37 °C in 5% CO<sub>2</sub> until 10 h after inoculation (that is, one virus life cycle). Cells were fixed at 10 h after inoculation for the quantification

of the percentage of infected cells using an immunofluorescence assay targeting SARS-CoV-2 N.

## In vitro primer elongation assay using RdRp core complex

Expression plasmids for SARS-CoV-2 nsp7, nsp8 and nsp12 were provided by the laboratory of S. Darst and E. Campbell. The expression and purification of nsp7, nsp8 and nsp12 were performed as previously described<sup>25</sup>. The transcription scaffold was prepared by annealing a 2:1:3 molar ratio of template strand RNA (5'-CUAUCCCCAUGUGAU UUUAAUAGCUUCUUAGGAGAAUGACGUAGCAUGCACGCG), <sup>32</sup>P labelled primer RNA (5'-CGCGUAGCAUGCUCGUACGUACGUUAUCUCCUAAGAACUA) and nontemplate DNA (5'-ATCACATGGGGATAG) at 95 °C for 5 min, and a slow cooldown to room temperature in elongation buffer (20 mM Tris pH 7.5, 40 mM KCl, 5 mM MgCl<sub>2</sub> and 5 mM DTT). The prepared scaffold was incubated with different concentrations of clofazimine or SL-11128 for 2 h at room temperature. SL-11128 has previously been reported to be effective against SARS-CoV-2<sup>15</sup>. The RdRp core complex was prepared by mixing a 1:3 molar ratio of nsp12 to nsp7 and nsp 8 and incubating for 20 min in ice. The scaffold was added to the RdRp complex and incubated for 20 min at 30 °C. To start the reaction, an equal volume of rNTP was added to the RdRp-scaffold complex. The final concentrations for assays were: 20 nM scaffold, 250 nM RdRp, 5–40 µM of clofazimine or SL-11128 and 10 µM rNTP. The reaction was quenched by mixing with loading buffer (90% formamide, 50 mM EDTA, 0.05% xylene cyanol and 0.05% bromophenol blue) at different time points. All samples were denatured at 95 °C for 20 min and analysed by 8% TBE and urea denaturing PAGE.

## In vitro helicase unwinding assay using nsp13

Helicase inhibition assays were performed as previously described<sup>37</sup>. Recombinant SARS-CoV-2 nsp13 was expressed in *Escherichia coli* BL21(DE3) cells and purified using the Ni<sup>2+</sup>-loaded HiTrap Chelating System (GE Healthcare) according to the manufacturer's instructions. The DNA oligomers FL-Cy3 oligonucleotide (5'-

TTTTTTTTTTTTTTTTTC  
GAGCACCGCTGCGGCTGCACC(Cy3)-3') and RL-BHQ oligonucleotide (5'-(BHQ2)GGTGCAGCCGCAGCGGTGCTG-3') were purchased from Metabion. RNA oligonucleotides, including RNA\_31/18-mer-Cy3 (5'-CGCAGUCUUCUCCUGGUGCUCGAAC AGUGAC(Cy3)-3') and RNA\_31/18-mer\_BHQ (5'-(BHQ2)GUCACUGUUCGAGCACCA-3'), were synthesized from IDT. Both oligonucleotides were annealed in the buffer of 20 mM Tris-HCl pH 8.0 and 150 mM NaCl. Assays were performed in buffer composed of 20 mM Tris-HCl buffer, pH 7.4, 150 mM NaCl, 0.1 mg ml<sup>-1</sup> BSA, 5 mM MgCl<sub>2</sub>, 5 mM TCEP, 5% glycerol and 10 nM helicase, followed by the addition of 0.5 µl of 100 mM ATP and 1.5 µl of oligonucleotide mixture to make the final concentration of FL-Cy3:RL-BHQ oligonucleotide and FL-Cy3:RL oligonucleotide 5 nM and 10 nM, respectively. Fluorescence ( $\lambda_{\text{ex}} = 550$  nm,  $\lambda_{\text{em}} = 620$  nm) was measured using SpectraMax iD3 Multi-Mode microplate reader to determine the extent of DNA–RNA duplex unwinding.

## Illustrations

Hamster illustrations in Fig. 3a and Extended Data Fig. 6a were created with BioRender software (<https://biorender.com/>).

## Reporting summary

Further information on research design is available in the [Nature Research Reporting Summary](#) linked to this paper.

## Data availability

The complete sequences of the SARS-CoV-2 HKU-001a and SARS-CoV-2 USA-WA1/2020 strains that we used were obtained through GenBank (accession numbers [MT230904](#) (HKU-001a), [MT246667](#) and [MN908947](#) (USA-WA1/2020)). The raw RNA-seq data discussed in this Article have been deposited in NCBI's Gene Expression Omnibus and are accessible through GEO series accession number [GSE162899](#). The hg38 reference genome was downloaded from UCSC database

(<https://hgdownload.soe.ucsc.edu/goldenPath/hg38/bigZips/>). The refseq gene annotation was retrieved from the UCSC Table Browser (<https://genome.ucsc.edu/cgi-bin/hgTables>). The NCBI SARS-CoV-2 reference genome (NC\_045512.2) was downloaded from the NCBI database (<https://www.ncbi.nlm.nih.gov/nuccore/1798174254>). The MesAur1.0 genome assembly (GCA\_000349665.1) and annotation were obtained from the Ensembl database ([https://asia.ensembl.org/Mesocricetus\\_auratus/Info/Index](https://asia.ensembl.org/Mesocricetus_auratus/Info/Index)). Any other relevant data are available from the corresponding authors upon reasonable request. [Source data](#) are provided with this paper.

## References

1. 1.

Peiris, J. S. et al. Coronavirus as a possible cause of severe acute respiratory syndrome. *Lancet* **361**, 1319–1325 (2003).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

2. 2.

Zaki, A. M., van Boheemen, S., Bestebroer, T. M., Osterhaus, A. D. & Fouchier, R. A. Isolation of a novel coronavirus from a man with pneumonia in Saudi Arabia. *N. Engl. J. Med.* **367**, 1814–1820 (2012).

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

3. 3.

Barry, V. C. et al. A new series of phenazines (rimino-compounds) with high antituberculosis activity. *Nature* **179**, 1013–1015 (1957).

[ADS](#) [CAS](#) [PubMed](#) [Google Scholar](#)

4. 4.

World Health Organization. Coronavirus Disease (COVID-19) Situation Reports. <https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports> (2020).

5. 5.

Chan, J. F. et al. A familial cluster of pneumonia associated with the 2019 novel coronavirus indicating person-to-person transmission: a study of a family cluster. *Lancet* **395**, 514–523 (2020).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

6. 6.

Hung, I. F. et al. SARS-CoV-2 shedding and seroconversion among passengers quarantined after disembarking a cruise ship: a case series. *Lancet Infect. Dis.* **20**, 1051–1060 (2020).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

7. 7.

Chan, J. F., To, K. K., Tse, H., Jin, D. Y. & Yuen, K. Y. Interspecies transmission and emergence of novel viruses: lessons from bats and birds. *Trends Microbiol.* **21**, 544–555 (2013).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

8. 8.

Chan, J. F. et al. Middle East respiratory syndrome coronavirus: another zoonotic betacoronavirus causing SARS-like disease. *Clin. Microbiol. Rev.* **28**, 465–522 (2015).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

9. 9.

Zumla, A., Chan, J. F., Azhar, E. I., Hui, D. S. & Yuen, K. Y. Coronaviruses – drug discovery and therapeutic options. *Nat. Rev. Drug Discov.* **15**, 327–347 (2016).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

10. 10.

Sheahan, T. P. et al. Broad-spectrum antiviral GS-5734 inhibits both epidemic and zoonotic coronaviruses. *Sci. Transl. Med.* **9**, eaal3653 (2017).

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

11. 11.

Beigel, J. H., Tomashek, K. M. & Dodd, L. E. Remdesivir for the treatment of Covid-19 – preliminary report. *Reply. N. Engl. J. Med.* **383**, 994 (2020).

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

12. 12.

Wang, Y. et al. Remdesivir in adults with severe COVID-19: a randomised, double-blind, placebo-controlled, multicentre trial. *Lancet* **395**, 1569–1578 (2020).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

13. 13.

Goldman, J. D. et al. Remdesivir for 5 or 10 days in patients with severe Covid-19. *N. Engl. J. Med.* **383**, 1827–1837 (2020).

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

14. 14.

Hung, I. F. et al. Triple combination of interferon beta-1b, lopinavir-ritonavir, and ribavirin in the treatment of patients admitted to hospital with COVID-19: an open-label, randomised, phase 2 trial. *Lancet* **395**, 1695–1704 (2020).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

15. 15.

Riva, L. et al. Discovery of SARS-CoV-2 antiviral drugs through large-scale compound repurposing. *Nature* **586**, 113–119 (2020).

[ADS](#) [CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

16. 16.

Gopal, M., Padayatchi, N., Metcalfe, J. Z. & O'Donnell, M. R. Systematic review of clofazimine for the treatment of drug-resistant tuberculosis. *Int. J. Tuberc. Lung Dis.* **17**, 1001–1007 (2013).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

17. 17.

Schaad-Lanyi, Z., Dieterle, W., Dubois, J. P., Theobald, W. & Vischer, W. Pharmacokinetics of clofazimine in healthy volunteers. *Int. J. Lepr. Other Mycobact. Dis.* **55**, 9–15 (1987).

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

18. 18.

Mansfield, R. E. Tissue concentrations of clofazimine (B663) in man. *Am. J. Trop. Med. Hyg.* **23**, 1116–1119 (1974).

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

19. 19.

Kissler, S. M., Tedijanto, C., Goldstein, E., Grad, Y. H. & Lipsitch, M. Projecting the transmission dynamics of SARS-CoV-2 through the postpandemic period. *Science* **368**, 860–868 (2020).

[ADS](#) [CAS](#) [PubMed](#) [Google Scholar](#)

20. 20.

Sharma, A. et al. Human iPSC-derived cardiomyocytes are susceptible to SARS-CoV-2 infection. *Cell Rep. Med.* **1**, 100052 (2020).

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

21. 21.

Yuan, S. et al. SREBP-dependent lipidomic reprogramming as a broad-spectrum antiviral target. *Nat. Commun.* **10**, 120 (2019).

[ADS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

22. 22.

Clausen, T. M. et al. SARS-CoV-2 infection depends on cellular heparan sulfate and ACE2. *Cell* **183**, 1043–1057 (2020).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

23. 23.

Lu, L. et al. Structure-based discovery of Middle East respiratory syndrome coronavirus fusion inhibitor. *Nat. Commun.* **5**, 3067 (2014).

[ADS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

24. 24.

Xia, S. et al. A pan-coronavirus fusion inhibitor targeting the HR1 domain of human coronavirus spike. *Sci. Adv.* **5**, eaav4580 (2019).

[ADS](#) [CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

25. 25.

Chen, J. et al. Structural basis for helicase-polymerase coupling in the SARS-CoV-2 replication–transcription complex. *Cell* **182**, 1560–1573 (2020).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

26. 26.

Chan, J. F.-W. et al. Simulation of the clinical and pathological manifestations of coronavirus disease 2019 (COVID-19) in golden Syrian hamster model: implications for disease pathogenesis and transmissibility. *Clin. Infect. Dis.* **71**, 2428–2446 (2020).

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

27. 27.

Nix, D. E. et al. Pharmacokinetics and relative bioavailability of clofazimine in relation to food, orange juice and antacid. *Tuberculosis* **84**, 365–373 (2004).

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

28. 28.

Herold, T. et al. Elevated levels of IL-6 and CRP predict the need for mechanical ventilation in COVID-19. *J. Allergy Clin. Immunol.* **146**, 128–136 (2020).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

29. 29.

Cholo, M. C., Steel, H. C., Fourie, P. B., Germishuizen, W. A. & Anderson, R. Clofazimine: current status and future prospects. *J.*

*Antimicrob. Chemother.* **67**, 290–298 (2012).

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

30. 30.

Hwang, T. J. et al. Safety and availability of clofazimine in the treatment of multidrug and extensively drug-resistant tuberculosis: analysis of published guidance and meta-analysis of cohort studies. *BMJ Open* **4**, e004143 (2014).

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

31. 31.

Falzon, D. et al. World Health Organization treatment guidelines for drug-resistant tuberculosis, 2016 update. *Eur. Respir. J.* **49**, 1602308 (2017).

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

32. 32.

Hadjadj, J. et al. Impaired type I interferon activity and inflammatory responses in severe COVID-19 patients. *Science* **369**, 718–724 (2020).

[ADS](#) [CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

33. 33.

Lucas, C. et al. Longitudinal analyses reveal immunological misfiring in severe COVID-19. *Nature* **584**, 463–469 (2020).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

34. 34.

Weng, Z. et al. A simple, cost-effective but highly efficient system for deriving ventricular cardiomyocytes from human pluripotent stem

cells. *Stem Cells Dev.* **23**, 1704–1716 (2014).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

35. 35.

Chan, J. F. et al. Development and evaluation of novel real-time reverse transcription–PCR assays with locked nucleic acid probes targeting leader sequences of human-pathogenic coronaviruses. *J. Clin. Microbiol.* **53**, 2722–2726 (2015).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

36. 36.

Chu, H. et al. Comparative replication and immune activation profiles of SARS-CoV-2 and SARS-CoV in human lungs: an ex vivo study with implications for the pathogenesis of COVID-19. *Clin. Infect. Dis.* **71**, 1400–1409 (2020).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

37. 37.

Yuan, S. et al. Metallodrug ranitidine bismuth citrate suppresses SARS-CoV-2 replication and relieves virus-associated pneumonia in Syrian hamsters. *Nat. Microbiol.* **5**, 1439–1448 (2020).

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

38. 38.

Kim, D. et al. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol.* **14**, R36 (2013).

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

39. 39.

Liao, Y., Smyth, G. K. & Shi, W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923–930 (2014).

[CAS](#) [Google Scholar](#)

40. 40.

Yu, G., Wang, L. G., Han, Y. & He, Q. Y. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS* **16**, 284–287 (2012).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

41. 41.

Zhou, Y. et al. Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nat. Commun.* **10**, 1523 (2019).

[ADS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

42. 42.

Hoffmann, M. et al. SARS-CoV-2 cell entry depends on ACE2 and TMPRSS2 and is blocked by a clinically proven protease inhibitor. *Cell* **181**, 271–280.e8 (2020).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

43. 43.

Xie, X. et al. An infectious cDNA clone of SARS-CoV-2. *Cell Host Microbe* **27**, 841–848.e3 (2020).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

44. 44.

Ianevski, A., Giri, A. K. & Aittokallio, T. SynergyFinder 2.0: visual analytics of multi-drug combination synergies. *Nucleic Acids Res.* **48**, W488–W493 (2020).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

[Download references](#)

## Acknowledgements

The authors acknowledge the assistance of the University of Hong Kong Li Ka Shing Faculty of Medicine Centre for PanorOmic Sciences. We thank S. Darst and E. Campbell for providing us plasmids of nsp7, nsp8 and nsp12; and K. Godula and R. Porell for heparin–BSA conjugate. This study was partly supported by funding to the University of Hong Kong: the Health and Medical Research Fund (grant no. COVID190121 and CID-HKU1-11), the Food and Health Bureau, The Government of the Hong Kong Special Administrative Region; the Innovation and Technology Fund, Innovation and Technology Commission, the Government of the Hong Kong Special Administrative Region; Sanming Project of Medicine in Shenzhen, China (SZSM201911014); and the High Level-Hospital Program, Health Commission of Guangdong Province, China; and donations from R. Y.-H. Yu and C. K.-W. Yu, the Shaw Foundation Hong Kong, M. S.-K. Tong, M. M.-Y. Tam Mak, Lee Wan Keung Charity Foundation Limited, Hong Kong Sanatorium & Hospital, the M. Hui, Hui Hoy and Chow Sin Lan Charity Fund Limited, the Chan Yin Chuen Memorial Charitable Foundation, M. M.-W. Lee, the Hong Kong Hainan Commercial Association South China Microbiology Research Fund, the Jessie & George Ho Charitable Foundation, Perfect Shape Medical Limited, Kai Chong Tong, Norman & Cecilia Yip Foundation, L. K.-M. Tse, Foo Oi Foundation Limited, B. H.-C. Lee, P. C. So and the Lo Ying Shek Chi Wai Foundation. This work was also supported by the grants to the Sanford Burnham Prebys Medical Discovery Institute: DoD: W81XWH-20-1-0270; DHIPC: U19 AI118610; NCI Cancer Center grant P30 CA030199-40; Fluomics/NOSI: U19 AI135972, as well as generous philanthropic donations from D. Ruch, and S. and J. Blair. The funding sources had no

role in the study design, data collection, analysis, interpretation or writing of the report.

## Author information

### Author notes

#### 1. Laura Riva

Present address: Calibr, Scripps Research, La Jolla, CA, USA

#### 2. These authors contributed equally: Shuofeng Yuan, Xin Yin, Xiangzhi Meng, Jasper Fuk-Woo Chan, Zi-Wei Ye

### Affiliations

#### 1. State Key Laboratory of Emerging Infectious Diseases, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Pokfulam, Hong Kong SAR, China

Shuofeng Yuan, Jasper Fuk-Woo Chan, Kong-Hung Sze, Anna Jinxia Zhang, Hin Chu, Kin-Hang Kok, Honglin Chen & Kwok-Yung Yuen

#### 2. Department of Microbiology, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Pokfulam, Hong Kong SAR, China

Shuofeng Yuan, Xiangzhi Meng, Jasper Fuk-Woo Chan, Zi-Wei Ye, Chris Chun-Yiu Chan, Pok-Man Lai, Chris Chung-Sing Chan, Vincent Kwok-Man Poon, Andrew Chak-Yiu Lee, Chun-Kit Yuen, Jianli Cao, Ronghui Liang, Kaiming Tang, Kong-Hung Sze, Anna Jinxia Zhang, Hin Chu, Kin-Hang Kok, Honglin Chen & Kwok-Yung Yuen

#### 3. Department of Clinical Microbiology and Infection Control, The University of Hong Kong–Shenzhen Hospital, Shenzhen, China

Shuofeng Yuan, Jasper Fuk-Woo Chan & Kwok-Yung Yuen

4. State Key Laboratory of Veterinary Biotechnology, Harbin Veterinary Research Institute, Chinese Academy of Agricultural Sciences, Harbin, China

Xin Yin, Yu-Yuan Zhang, Yan-Dong Tang & Xue-Hui Cai

5. Immunity and Pathogenesis Program, Infectious and Inflammatory Disease Center, Sanford Burnham Prebys Medical Discovery Institute, La Jolla, CA, USA

Xin Yin, Laura Riva, Lars Pache, Naoko Matsunaga, Yuan Pu & Sumit K. Chanda

6. Hainan Medical University–The University of Hong Kong Joint Laboratory of Tropical Infectious Diseases, Hainan Medical University and The University of Hong Kong, Pokfulam, Hong Kong SAR, China

Jasper Fuk-Woo Chan & Kwok-Yung Yuen

7. School of Biomedical Sciences, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Pokfulam, Hong Kong SAR, China

Li Sheng, Wan Xu, Chit-Ying Lau, Dong-Yan Jin & Ren Sun

8. Department of Molecular and Medical Pharmacology, David Geffen School of Medicine, University of California Los Angeles, Los Angeles, CA, USA

Li Sheng, Yushen Du & Ren Sun

9. Department of Surgery, The University of Hong Kong, Queen Mary Hospital, Pokfulam, Hong Kong SAR, China

Ko-Yung Sit & Wing-Kuk Au

10. Department of Chemistry, State Key Laboratory of Synthetic Chemistry, CAS–HKU Joint Laboratory of Metallomics on Health and Environment, The University of Hong Kong, Pokfulam, Hong Kong SAR, China

Runming Wang & Hongzhe Sun

11. Department of Cellular and Molecular Medicine, University of California San Diego, La Jolla, CA, USA  
Thomas Mandel Clausen, Jessica Pihl, Juntaek Oh, Dong Wang & Jeffrey D. Esko
12. Copenhagen Center for Glycomics, Department of Molecular and Cellular Medicine, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark  
Thomas Mandel Clausen
13. Department of Immunology and Microbiology, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark  
Jessica Pihl
14. Division of Pharmaceutical Sciences, Skaggs School of Pharmacy and Pharmaceutical Sciences, University of California San Diego, La Jolla, CA, USA  
Juntaek Oh & Dong Wang
15. Glycobiology Research and Training Center, University of California San Diego, La Jolla, CA, USA  
Jeffrey D. Esko
16. Department of Medicine, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Pokfulam, Hong Kong SAR, China  
Ivan Fan-Ngai Hung
17. Dr Li Dak-Sum Research Centre, The University of Hong Kong, Pokfulam, Hong Kong SAR, China  
Ronald Adolphus Li

## Authors

1. Shuofeng Yuan

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

2. Xin Yin

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

3. Xiangzhi Meng

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

4. Jasper Fuk-Woo Chan

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

5. Zi-Wei Ye

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

6. Laura Riva

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

7. Lars Pache

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

8. Chris Chun-Yiu Chan

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

9. Pok-Man Lai

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

10. Chris Chung-Sing Chan

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

11. Vincent Kwok-Man Poon

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

12. Andrew Chak-Yiu Lee

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

13. Naoko Matsunaga

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

14. Yuan Pu

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

15. Chun-Kit Yuen

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

16. Jianli Cao

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

17. Ronghui Liang

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

18. Kaiming Tang

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

19. Li Sheng

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

20. Yushen Du

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

21. Wan Xu

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

22. Chit-Ying Lau

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

23. Ko-Yung Sit

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

24. Wing-Kuk Au

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

25. Runming Wang

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

26. Yu-Yuan Zhang

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

27. Yan-Dong Tang

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

28. Thomas Mandel Clausen

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

29. Jessica Pihl

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

30. Juntaek Oh

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

31. Kong-Hung Sze

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

32. Anna Jinxia Zhang

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

33. Hin Chu

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

34. Kin-Hang Kok

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

35. Dong Wang

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

36. Xue-Hui Cai

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

37. Jeffrey D. Esko

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

38. Ivan Fan-Ngai Hung

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

39. Ronald Adolphus Li

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

40. Honglin Chen

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

41. Hongzhe Sun

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

42. Dong-Yan Jin

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

43. Ren Sun

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

44. Sumit K. Chanda

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

45. Kwok-Yung Yuen

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

## Contributions

S.Y., X.Y., X.M., Z.-W.Y., L.R., N.M., L.P., Y.P., R.A.L., C.C.-Y.C., P.-M.L., J.C., R.L., Y.-Y.Z., Y.-D.T. and W.X. designed and/or performed experiments. S.Y., X.Y., X.M., Z.-W.Y., J.F.-W.C., L.P., Y.D., K.-H.S. and H. Chu analysed data. S.Y., X.Y., Z.-W.Y., C.-K.Y., K.-H.K., K.-Y.S, W.-K.A., R.W., Y.-Y.Z., Y.-D.T., C.-Y.L., T.M.C., J.P., J.O. and L.S. generated critical reagents. K.-Y.Y., S.K.C., R.S., S.Y., X.Y., X.M. and L.P. wrote the manuscript. S.Y., J.F.-W.C., C.C.-S.C., V.K.-M.P. and K.T. provided the hamster data. A.C.-Y.L, A.J.Z., S.Y. and J.F.-W.C. performed histopathological evaluation. S.K.C., R.S., I.F.-N.H., R.A.L., H. Chen, H.S.,

D.W., X.-H.C., J.D.E. and D.-Y.J. provided conceptual advice and troubleshooting. K.-Y.Y., S.K.C. and R.S. oversaw the conception and supervised the study. K.-Y.Y. and S.K.C. provided the grant support.

## **Corresponding authors**

Correspondence to [Ren Sun](#) or [Sumit K. Chanda](#) or [Kwok-Yung Yuen](#).

## **Ethics declarations**

## Competing interests

J.F.-W.C. has received travel grants from Pfizer Corporation Hong Kong and Astellas Pharma Hong Kong Corporation Limited, and was an invited speaker for Gilead Sciences Hong Kong Limited and Luminex Corporation. S.K.C. is an inventor on a patent application on repurposed antiviral compounds for SARS-CoV-2 owned by Scripps Research and Sanford Burnham Prebys. US Patent Application Serial No. 63/010630, entitled ‘Methods and Compositions for Antiviral Treatment’ relates to aspects of this work and was filed on 15 April 2020. The corresponding authors had full access to all the data in the study and had final responsibility for the decision to submit for publication. The other authors declare no competing interests.

## Additional information

**Peer review information** *Nature* thanks the anonymous reviewer(s) for their contribution to the peer review of this work. Peer review reports are available.

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

## Extended data figures and tables

### [Extended Data Fig. 1 Anti-MERS-CoV activity and cytotoxicity measurement of clofazimine in matching cells.](#)

**a**, Clofazimine inhibited MERS-CoV replication in a dose-dependent manner. EC<sub>50</sub> was achieved by plaque reduction assay and plotted using logistic nonlinear regression model (GraphPad Prism 7). Data are mean ± s.d., n = 3 biologically independent samples. **b**, MERS-CoV-infected Huh7 cells (MOI of 0.01) were treated with clofazimine (5 µM) or were not treated. Top, immunofluorescence staining of MERS-CoV-N antigen (green), and Huh7 cell nucleus (blue). Scale bars, 20 µm.

Representative images selected from a pool images captured in two independent experiments. Bottom, MERS-CoV-N-positive cells quantified by flow cytometry. The experiments were performed twice and representative quantifications are shown. **c**, The cell viability was determined using CellTiter-Glo assays and in the absence of virus infection. The drug-incubation time in the cytotoxicity assay was consistent with that in the antiviral assay: for example, at 24 h after treatment for Huh7 cells, primary human small-airway epithelial cells (HSAEpC) and human embryonic stem-cell-derived cardiomyocytes (CM); at 48 h after treatment for Vero E6 cells; and at 72 h after treatment for BSC1 and human embryonic lung fibroblasts (HELF). Data are mean  $\pm$  s.d.,  $n = 3$  biological replicates. The experiment was repeated twice for confirmation.

### **Extended Data Fig. 2 Exploration of possible effects of clofazimine on virus entry and replication.**

**a**, Time-of-addition assay. SARS-CoV-2-infected VeroE6 cells were incubated with clofazimine and at the time points indicated. Infection at 10 h after infection was quantified by immunostaining for N. Data are normalized to the DMSO-treated cells, and are mean  $\pm$  s.d.,  $n = 6$  independent experiments. One-way ANOVA followed by Tukey's post hoc test. **b**, Clofazimine has no effect on ACE2 and DPP4 expression. Caco-2 cells were treated with clofazimine for 16 h before collection for western blotting analysis. The expression of ACE2 and DPP4 were determined using anti-ACE2 antibody (Abcam; ab108252; 1:1,000) and anti-DPP4 antibody (Cell Signaling; 67138T; 1:1,000), respectively. For gel source data, see Supplementary Fig. 1. The experiment was repeated twice for confirmation. **c**, Clofazimine has no inhibition on the binding between either ACE2 or heparin and SARS-CoV-2 S, which are two critical cellular components for viral attachment and infection. Clofazimine was titrated as the indicated concentrations. Dashed line represents binding without inhibitor (that is, 0  $\mu$ M). Data are mean  $\pm$  s.d.,  $n = 3$  biologically independent samples. **d**, Quantification on the basis of GFP positive area using ImageJ software (corresponding to Fig. 2b). Error bars are s.e.m.,  $n = 5$  randomly selected images. One-way ANOVA followed by Dunnett's post test. **e**, Scheme of in vitro-transcribed viral RNA replication assay. Vero E6 cells were electroporated with in vitro-transcribed viral RNA. At 2

h after seeding, the cells were treated with the drug. Negative-stranded RNA was then quantified at 12 h after electroporation. **f**, GFP mRNA was used as a negative control and its translation, determined at 24 h after electroporation, was not affected by remdesivir or clofazimine. Error bars are s.e.m.,  $n = 3$  independent experiments. n.s., not significant ( $P > 0.05$ , when compared with the 0- $\mu\text{M}$  group by two-tailed Student's *t*-test). **g**, Clofazimine shows a marginal effect against main protease ( $M^{\text{pro}}$ ) and papain-like protease ( $PL^{\text{pro}}$ ) activity. The activity of purified SARS-CoV-2  $M^{\text{pro}}$  and SARS-CoV-2  $PL^{\text{pro}}$  enzymes was measured after adding their respective peptide substrates. Enzyme activity in the absence (100% activity) and presence of clofazimine were calculated. Data are mean  $\pm$  s.d.,  $n = 3$  independent experiments. **h**, Clofazimine shows no inhibition on the polymerase activity of nsp7–nsp8–nsp12–RdRp complex. The scaffold used in this in vitro transcription inhibition assay is listed in the top panel. Inhibition effect of clofazimine or SL-11128 on the RdRp core complex was analysed by a primer elongation assay. SL-11128, a SARS-CoV-2 inhibitor<sup>15</sup>, shows some inhibition effect over time (0, 15 s, 1 min, 5 min and 20 min), whereas that of clofazimine was minor (ranging from around 5 to 40  $\mu\text{M}$ ). All of the reactions were performed at 30 °C. All the above experiments were repeated twice for confirmation.

### Extended Data Fig. 3 Transcriptional analysis of clofazimine treatment.

**a**, Caco-2 cells were infected with SARS-CoV-2 at an MOI of 0.1 before being treated with clofazimine, remdesivir and DMSO (0.1%) at the indicated concentrations. Cell culture supernatant was collected at 48 h after infection, and was subjected to viral load determination. Two-tailed Student's *t*-test, compared with the DMSO group. \* $P < 0.05$ , \*\* $P < 0.01$ . Data are mean  $\pm$  s.d.,  $n = 3$  independent experiments. **b**, Timeline of the transcriptomic study (MOI = 4). h, hours after drug treatment; hpi, hours post virus infection; CFZ, clofazimine. **c**, Principal components analysis of RNA-seq dataset after ‘reads per kilobase of transcript per million reads mapped’ normalization on each gene expression level. Each dot represents one sample. The percentage labelled on the *x* or *y* axis represents the proportion of variance explained with each principal component (PC). **d**,

Patterns of transcription levels across all samples. The genes that were significantly and differentially expressed (fold change  $>2$  or  $<0.5$ , FDR  $< 0.05$ ) between 6 h after infection and mock are shown. Conditions include 3 h and 6 h after infection for Caco-2 cells (MOI = 4) with or without clofazimine treatment. Genes were clustered using the *K*-means method. **e**, Heat map of 197 transcription factors regulated by clofazimine treatment without infection (left), and known interactions among these transcription factors (right). **f**, Heat map of the genes enriched in MAPK signalling, TNF signalling, interleukin signalling or cytokine–cytokine receptor interaction. These genes are upregulated (fold change  $>2$ , FDR  $< 0.05$ ) by either 6 h after treatment with clofazimine (without infection) or at 6 h after infection with clofazimine treatment, compared to mock infection.

**Extended Data Fig. 4 Transcriptional comparison of clofazimine-treated cells with or without SARS-CoV-2 infection.**

**a**, Top-enriched pathways of significantly upregulated genes (FDR  $< 0.05$ , fold change  $> 2$ ) compared between clofazimine treatment alone at 6 h versus mock-infected cells (6 h CFZ versus mock) and at 6 h after infection under clofazimine treatment versus mock (6 h CFZ versus mock). Pathway analysis was performed by Metascape. \*Innate-immunity-related pathways. **b**, Network of enriched terms represented as pie charts. Pies are colour-coded on the basis of the identities of the gene lists. ‘6 h. clofazimine vs mock’, upregulated genes by clofazimine treatment without infection at 6 h, compared with mock; ‘6 hpi. clofazimine vs mock’, upregulated genes by clofazimine treatment at 6 h after infection, and compared with mock.

**Extended Data Fig. 5 Prophylactic and therapeutic administration of clofazimine reduced viral load in the lungs without compromising humoral response in hamsters.**

Experimental design is shown in Fig. 3a. **a**, Viral yield in lung tissue from hamsters, after prophylactic ( $n = 5$  hamsters per group) or therapeutic treatment ( $n = 11$  hamsters (each remdesivir and clofazimine group) and  $n = 13$  hamsters (vehicle group)), collected at 4 days after infection and

titrated by RT–qPCR assays. Data are mean  $\pm$  s.d. Two-tailed Student's *t*-test for prophylaxis groups and one-way ANOVA followed by Dunnett's post test for therapeutic groups, comparing with the vehicle group (black symbols). \*\*\* $P$  < 0.001, \*\* $P$  < 0.01. **b**, Hamsters exhibited a normal humoral immune response after SARS-CoV-2 infection and clofazimine treatment. The sera were serially diluted before adding to the N-coated ELISA plate ( $n$  = 3 hamsters per group). [Source data](#)

### Extended Data Fig. 6 Transcriptional analysis of hamster lung tissues with clofazimine administration.

**a**, Experimental design. Tissue samples were collected at the indicated time points. **b**, Gene Ontology Biological Process analysis results for upregulated genes, comparing prophylactic clofazimine administration with its corresponding vehicle controls. **c**, RNA expression (reads per kilobase of transcript per million mapped reads) of the 13 genes enriched in the 'leukocyte differentiation' category of the Gene Ontology Biological Process analysis. These genes are upregulated (fold change > 1.5,  $P$  value < 0.01) in the prophylactic clofazimine group versus vehicle controls. \*MHC class-II molecules. ↑Transcription factors upregulated by clofazimine in both Caco-2 cells and hamster lung tissues. Data are mean  $\pm$  s.e.m.,  $n$  = 3 hamsters per group. **d**, Heat map of immune-response-related genes in lungs of uninfected or infected hamsters administered prophylactic clofazimine or vehicle controls. **e**, Gene Ontology Biological Process analysis results for upregulated genes, comparing clofazimine- and vehicle-treated hamster lungs without virus infection. [Source data](#)

### Extended Data Fig. 7 Histological analysis lung pathology in each group.

**a**, Representative images of H&E-stained lung tissue section from hamsters treated as indicated. Numbered circled areas are shown in magnified images to the right, illustrating the severity of (1) bronchiolar and/or peribronchiolar cell death; (2) alveoli destruction and/or alveolar infiltration; and (3) blood vessel and perivascular infiltration. Black arrows indicate sites of infiltration. These representative images were selected from

a pool of 15 images captured in 3 randomly selected hamsters per group. Scale bars, 200 µm. **b**, Semiquantitative histology scores were given to each lung tissue by grading the severity of damage in bronchioles, alveoli and blood vessels and accumulating the total scores. For bronchioles: 0 = normal structure; 1 = mild peribronchiolar infiltration; 2 = peribronchiolar infiltration plus epithelial cell death; and 3 = score 2 plus intrabronchiolar wall infiltration and epithelium desquamation. For alveoli: 0 = normal structure; 1 = alveolar wall thickening and congestion; 2 = focal alveolar space infiltration or exudation; and 3 = diffuse alveolar space infiltration or exudation or haemorrhage. For blood vessels: 0 = normal structure; 1 = mild perivascular oedema or infiltration; 2 = vessel wall infiltration; and 3 = severe endothelium infiltration. Data are mean ± s.d., 3 randomly selected slides for each group. Unpaired two-tailed Student's *t*-test between the two prophylactic groups. One-way ANOVA followed by Dunnett's post test for the therapeutic groups. \*\**P* < 0.01, \*\*\**P* < 0.001, compared to the vehicle-control group. The histological score of mock infection was set as zero. [Source data](#)

### **Extended Data Fig. 8 Clofazimine exhibits antiviral synergy with remdesivir in vitro.**

**a**, Remdesivir at the indicated doses was combined with clofazimine at the indicated doses or a negative control (DMSO), and antiviral dose–response relationships were determined. Vero E6 cells were pre-treated for 16 h with increasing concentrations of the indicated compound and then infected with SARS-CoV-2 at a MOI of 0.01. Thirty hours after infection, the infected cells were analysed by immunofluorescence imaging. For each condition, the percentage of infection was calculated as the ratio of the number of infected cells stained for SARS-CoV-2 N to the number of cells stained with DAPI. Data are normalized to mean values for DMSO-treated wells, and are mean ± s.e.m. (*n* = 4 biological repeats over 2 independent experiments). **b**, Topographical two-dimensional map of synergy scores determined in synergy finder<sup>44</sup>. The colour gradient indicates the synergy score (red, highest score). *x* axis, remdesivir up to 10 µM, *y* axis, clofazimine up to 10 µM. **c**, Dose–response analysis of remdesivir alone (black) and in combination with 0.15625 µM (blue) or 0.625 µM (red) clofazimine. The observed compound activities are represented by solid

lines, and the predicted additive combinatorial activities are indicated by dashed lines. The dotted black line denotes 90% inhibition of infection. Data are normalized to mean values for DMSO-treated wells, and are mean  $\pm$  s.e.m. ( $n = 4$  biological repeats over 2 independent experiments). **d**, Counting of cell numbers in each drug combination, as indicated. Mean  $\pm$  s.d.,  $n = 4$  biologically independent samples. The experiments were repeated twice for confirmation. [Source data](#)

### **Extended Data Fig. 9 Clofazimine exhibits antiviral synergy with remdesivir in hamsters.**

Experiments were performed as shown in Fig. 4b. **a**, Representative image of infected cells by immunofluorescence staining in lung at 4 days after infection. SARS-CoV-2 N expression (green) is shown in diffuse alveolar areas (thick white arrow) and in the focal bronchiolar epithelial cells (thin white arrow) of the vehicle-treated hamster lungs, whereas standard and low-dose remdesivir (Rem) groups as well as the clofazimine group (Clo) exhibit reduced N expression. Combinatorial therapy restricts the virus replication within the entry gate of lung infection (that is, bronchiolar epithelial cells). These representative images were selected from a pool of 15 images captured in 3 randomly selected hamsters per group. **b**, Representative image of infected cells by immunofluorescence staining in nasal turbinate at 4 days after infection. SARS-CoV-2 N (NP) (green) and cell nuclei (blue) were stained. Scale bars, 200  $\mu$ m. These representative images were selected from a pool of 15 images captured in 5 hamsters per group. **c**, N-positive cells per 50 $\times$  field per the nasal turbinate section of a hamster. One-way ANOVA followed by Dunnett's post test and compared with vehicle ( $n = 5$  hamsters per group). \*\* $P < 0.01$ , n.s., not significant. [Source data](#)

## **Supplementary information**

### **Supplementary Information**

This file includes Supplementary Figures 1 and 2.

## Reporting Summary

### Supplementary Table 1

List of up-regulated genes comparing SARS-CoV-2 infected Caco-2 cells at 6 hpi versus un-infected cells. FDR<0.05, fold change>2 or <0.5 compared with mock.

### Supplementary Table 2

List of down-regulated genes comparing SARS-CoV-2 infected Caco-2 cells at 6 hpi versus un-infected cells. FDR<0.05, fold change>2 or <0.5 compared with mock.

### Supplementary Table 3

List of up-regulated genes in hamster lung tissues comparing prophylactic administration of Clofazimine versus vehicle control group at 4 dpi. Fold change >1.5, p<0.05 in prophylactic clofazimine administration group compared with vehicle controls at 4 dpi.

## Peer Review File

## **Source data**

### Source Data Fig. 3

### Source Data Fig. 4

### Source Data Extended Data Fig. 5

### Source Data Extended Data Fig. 6

### Source Data Extended Data Fig. 7

[\*\*Source Data Extended Data Fig. 8\*\*](#)

[\*\*Source Data Extended Data Fig. 9\*\*](#)

## Rights and permissions

[Reprints and Permissions](#)

## About this article



Check for  
updates

## Cite this article

Yuan, S., Yin, X., Meng, X. *et al.* Clofazimine broadly inhibits coronaviruses including SARS-CoV-2. *Nature* **593**, 418–423 (2021).  
<https://doi.org/10.1038/s41586-021-03431-4>

[Download citation](#)

- Received: 30 September 2020
- Accepted: 08 March 2021
- Published: 16 March 2021
- Issue Date: 20 May 2021
- DOI: <https://doi.org/10.1038/s41586-021-03431-4>

## Further reading

- [The Implementation of Mass-Vaccination against SARS-CoV-2: A Systematic Review of Existing Strategies and Guidelines](#)

- Tasnim Hasan
- , Justin Beardsley
- , Ben J. Marais
- , Thu Anh Nguyen
- & Greg J. Fox

*Vaccines* (2021)

## Comments

By submitting a comment you agree to abide by our [Terms](#) and [Community Guidelines](#). If you find something abusive or that does not comply with our terms or guidelines please flag it as inappropriate.

[Download PDF](#)

Advertisement

---

This article was downloaded by **calibre** from <https://www.nature.com/articles/s41586-021-03431-4>

Bispecific IgG neutralizes SARS-CoV-2 variants and prevents escape in mice

[Download PDF](#)

- Article
- [Published: 25 March 2021](#)

# Bispecific IgG neutralizes SARS-CoV-2 variants and prevents escape in mice

- [Raoul De Gasparo<sup>1 na1</sup>](#),
- [Mattia Pedotti<sup>1 na1</sup>](#),
- [Luca Simonelli<sup>1</sup>](#),
- [Petr Nickl ORCID: orcid.org/0000-0002-4289-0502<sup>2</sup>](#),
- [Frauke Muecksch ORCID: orcid.org/0000-0002-0132-5101<sup>3</sup>](#),
- [Irene Cassaniti<sup>4</sup>](#),
- [Elena Percivalle<sup>4</sup>](#),
- [Julio C. C. Lorenzi<sup>5</sup>](#),
- [Federica Mazzola<sup>1</sup>](#),
- [Davide Magri ORCID: orcid.org/0000-0003-3790-5332<sup>6</sup>](#),
- [Tereza Michalcikova<sup>2</sup>](#),
- [Jan Haviernik<sup>7</sup>](#),
- [Vaclav Honig ORCID: orcid.org/0000-0003-0469-4604<sup>7,8</sup>](#),
- [Blanka Mrazkova<sup>2</sup>](#),
- [Natalie Polakova<sup>2</sup>](#),
- [Andrea Fortova<sup>7</sup>](#),
- [Jolana Tureckova<sup>2</sup>](#),
- [Veronika Iatsiuk<sup>2</sup>](#),
- [Salvatore Di Girolamo<sup>1</sup>](#),
- [Martin Palus<sup>7,8</sup>](#),
- [Dagmar Zudova<sup>2</sup>](#),
- [Petr Bednar<sup>7,9</sup>](#),
- [Ivana Bukova<sup>2</sup>](#),
- [Filippo Bianchini<sup>1</sup>](#),
- [Dora Mehn<sup>6</sup>](#),
- [Radim Nencka<sup>10</sup>](#),
- [Petra Strakova<sup>7</sup>](#),

- [Oto Pavlis<sup>11</sup>](#),
- [Jan Rozman](#) ORCID: [orcid.org/0000-0002-8035-8904<sup>2</sup>](https://orcid.org/0000-0002-8035-8904),
- [Sabrina Gioria<sup>6</sup>](#),
- [Josè Camilla Sammartino<sup>4</sup>](#),
- [Federica Giardina<sup>4</sup>](#),
- [Stefano Gaiarsa<sup>4</sup>](#),
- [Qiang Pan-Hammarström](#) ORCID: [orcid.org/0000-0003-1990-8804<sup>12</sup>](https://orcid.org/0000-0003-1990-8804),
- [Christopher O. Barnes<sup>13</sup>](#),
- [Pamela J. Bjorkman](#) ORCID: [orcid.org/0000-0002-2277-3990<sup>13</sup>](https://orcid.org/0000-0002-2277-3990),
- [Luigi Calzolai](#) ORCID: [orcid.org/0000-0002-8474-7974<sup>6</sup>](https://orcid.org/0000-0002-8474-7974),
- [Antonio Piralla](#) ORCID: [orcid.org/0000-0002-6062-2579<sup>4</sup>](https://orcid.org/0000-0002-6062-2579),
- [Fausto Baldanti<sup>4,14</sup>](#),
- [Michel C. Nussenzweig<sup>5,15</sup>](#),
- [Paul D. Bieniasz](#) ORCID: [orcid.org/0000-0002-2368-3719<sup>3,15</sup>](https://orcid.org/0000-0002-2368-3719),
- [Theodora Hatzioannou<sup>3</sup>](#),
- [Jan Prochazka<sup>2</sup>](#),
- [Radislav Sedlacek](#) ORCID: [orcid.org/0000-0002-3352-392X<sup>2</sup>](https://orcid.org/0000-0002-3352-392X),
- [Davide F. Robbiani](#) ORCID: [orcid.org/0000-0001-7379-3484<sup>1</sup>](https://orcid.org/0000-0001-7379-3484),
- [Daniel Ruzek](#) ORCID: [orcid.org/0000-0003-4655-2380<sup>7,8</sup>](https://orcid.org/0000-0003-4655-2380) &
- [Luca Varani](#) ORCID: [orcid.org/0000-0002-0963-0987<sup>1</sup>](https://orcid.org/0000-0002-0963-0987)

[Nature](#) volume **593**, pages 424–428 (2021) [Cite this article](#)

- 25k Accesses
- 2 Citations
- 282 Altmetric
- [Metrics details](#)

## Subjects

- [Antibody therapy](#)
- [Immunization](#)
- [Protein design](#)
- [SARS-CoV-2](#)
- [Viral infection](#)

## Abstract

Neutralizing antibodies that target the receptor-binding domain (RBD) of the SARS-CoV-2 spike protein are among the most promising approaches against COVID-19<sup>1,2</sup>. A bispecific IgG1-like molecule (CoV-X2) has been developed on the basis of C121 and C135, two antibodies derived from donors who had recovered from COVID-19<sup>3</sup>. Here we show that CoV-X2 simultaneously binds two independent sites on the RBD and, unlike its parental antibodies, prevents detectable spike binding to the cellular receptor of the virus, angiotensin-converting enzyme 2 (ACE2). Furthermore, CoV-X2 neutralizes wild-type SARS-CoV-2 and its variants of concern, as well as escape mutants generated by the parental monoclonal antibodies. We also found that in a mouse model of SARS-CoV-2 infection with lung inflammation, CoV-X2 protects mice from disease and suppresses viral escape. Thus, the simultaneous targeting of non-overlapping RBD epitopes by IgG-like bispecific antibodies is feasible and effective, and combines the advantages of antibody cocktails with those of single-molecule approaches.

[Download PDF](#)

## Main

The COVID-19 pandemic has prompted substantial efforts to develop effective countermeasures against SARS-CoV-2. Preclinical data and phase-III clinical studies indicate that monoclonal antibodies could be effectively deployed for prevention or treatment during the viral symptoms phase of the disease<sup>1,2</sup>. Cocktails of two or more monoclonal antibodies are preferred over a single antibody as these cocktails result in increased efficacy and the prevention of viral escape. However, this approach requires increased manufacturing costs and volumes, which are problematic at a time when the supply chain is under pressure to meet the high demand for COVID-19 therapeutic agents, vaccines and biologics in general<sup>4</sup>. Cocktails also complicate formulation<sup>5,6</sup> and hinder strategies such as antibody delivery by viral vectors or by nonvectored nucleic acids<sup>7,8</sup>. One alternative is to use multispecific antibodies, which have the advantages of cocktails and single-molecule strategies.

To this end, we used structural information<sup>9</sup> and computational simulations to design bispecific antibodies that would simultaneously bind to (i) independent sites on the same RBD and (ii) distinct RBDs on a spike (S) trimer. We evaluated several designs using atomistic molecular dynamics simulations, and produced four constructs: of these, CoV-X2 was the most potent neutralizer of SARS-CoV-2 pseudovirus, and had a half-maximal inhibitory concentration ( $IC_{50}$ ) of 0.04 nM (5.8 ng ml<sup>-1</sup>) (Extended Data Fig. 1). CoV-X2 is a human-derived IgG1-like bispecific antibody in the CrossMAb format<sup>10</sup> that is the result of the combination of the Fragment antigen binding (Fab) of the monoclonal antibodies C121 and C135, which are two potent

neutralizers of SARS-CoV-2<sup>3</sup>. Structural predictions showed that CoV-X2—but not its parental monoclonal antibodies—can bind bivalently to all RBD conformations on the S trimer, which prevents the binding of ACE2 receptor<sup>11</sup> (Fig. 1a, Extended Data Fig. 2).

**Fig. 1: Biochemical and in vitro neutralizing properties of CoV-X2 are superior to those of its parental monoclonal antibodies.**

 figure1

**a**, Computational simulations predict bivalent binding of CoV-X2 to all three RBDs on the S trimer (Extended Data Fig. 2). Green and blue, C121 and C135 moieties, respectively; yellow and orange, RBDs. **b, c**, Surface plasmon resonance (SPR) demonstrates that both arms of CoV-X2 are functional. In **b**, immobilized RBD complexed with C121 (left) or C135 (right) (first antibody) binds to CoV-X2 (second antibody). In **c**, RBD–CoV-X2 prevents binding by C121 (left) or C135 (right) single

monoclonal antibodies. Paler colours denote controls (second antibody only). **d**, Both arms of CoV-X2 bind simultaneously to the RBD, as avidity is retained at decreasing RBD concentrations. Top and middle, representative SPR traces indicating the dissociations of Fab (left), IgG (centre) or CoV-X2 (right) binding to RBD immobilized at 150 nM (top) or 15 nM (middle) on the SPR chip (Extended Data Fig. 6). Bottom, plots of normalized  $k_a$  (left) and  $k_d$  (right) obtained with different concentrations of immobilized RBD. Increasing normalized  $k_d$  values indicate loss of avidity. Solid lines, IgG; dotted lines, Fab. **e, f**, CoV-X2 prevents ACE2 binding to S trimer in an ELISA. ACE2 binding to antibody–S-trimer complexes measured with increasing concentration of indicated antibody and constant ACE2 (**e**), or at constant antibody concentration with increasing ACE2 (**f**). Mean of two replicates shown. **g**, CoV-X2 neutralizes SARS-CoV-2 pseudovirus and escape mutants generated by its parental monoclonal antibodies. Normalized relative luminescence (RLU) for cell lysates after infection with NanoLuc-expressing SARS-CoV-2 pseudovirus in the presence of increasing concentrations of antibodies. Wild-type SARS-CoV-2 pseudovirus (top left) is shown alongside an escape mutant generated in the presence of C135 (R346S; top right), or two escape mutants generated in the presence of C121 (E484G (bottom left) and Q493R (bottom right))<sup>14</sup>. Dashed lines, parental Fabs; solid lines, IgG. Mean of two independent experiments with two replicates each. **h**, Neutralization of SARS-CoV-2 isolates with sequences corresponding to wild-type virus (which was first isolated in China), and the B.1 (D614G), B.1.1.7, P.1 and B.1.351 variants. Mean of three experiments with s.d. RBD residues mutated in the variants are indicated in the table and as red spheres on the S trimer structure, on which the epitopes of C135 (blue) and C121 (green) are shown.

[Full size image](#)

CoV-X2 bound at a low nanomolar affinity to the RBD and S trimer of wild-type SARS-CoV-2 and to those of several naturally occurring SARS-CoV-2 variants, including B.1 (which contains D614G, in the S), B.1.1.7 (which contains N501Y, in the RBD) and B.1.351 (which contains K417N/E484K/N501Y, in the RBD)<sup>12,13</sup>, as well as to escape mutants generated by the parental monoclonal antibodies<sup>14</sup> (Extended Data Figs. 3–5).

CoV-X2 also bound to preformed C121–RBD and C135–RBD complexes, which confirmed that both of its arms are functional (Fig. 1b, c). Next, we used an avidity assay by surface plasmon resonance to experimentally confirm the computational prediction that CoV-X2 can simultaneously engage two sites on the same RBD (Fig. 1d, Extended Data Fig. 6, Methods). Avidity occurs when IgG binds bivalently to antigens, which results in slower dissociation rates ( $k_d$ ) (Extended Data Fig. 6a). Accordingly, C121 and C135 IgG showed avidity at high concentrations of antigen, owing to the intermolecular binding of adjacent RBDs; at lower concentrations of

antigen, the  $k_d$  was faster as intermolecular binding was prevented by the increased distance between RBD molecules, which resulted in loss of avidity. Intramolecular avidity is not possible for C121 and C135 because a single epitope is available on each RBD molecule. By contrast, CoV-X2 maintained avidity even at low concentrations of antigen, which indicates that it undergoes bivalent, intramolecular binding (Fig. 1d, Extended Data Fig. 6). We then performed enzyme-linked immunosorbent assays (ELISAs) to evaluate the ability of CoV-X2 to inhibit the binding of recombinant ACE2 to the S trimer (Fig. 1e,f). Consistent with the structural information<sup>9</sup>, C135 did not affect the interaction between ACE2 and S. C121, which occupies the ACE2-binding site on the RBD, prevented ACE2 binding only partially. By contrast, ACE2 binding was not detected in the presence of CoV-X2, which suggests a synergistic effect by the two moieties that comprise the bispecific antibody.

To assess the neutralizing ability of CoV-X2 in vitro, we first used SARS-CoV-2 pseudoviruses<sup>15</sup>. The bispecific antibody neutralized pseudovirus that carries wild-type SARS-CoV-2 S at sub-nanomolar concentrations ( $IC_{50} = 0.04\text{ nM}$  ( $5.8\text{ ng ml}^{-1}$ ) and 90% inhibitory concentration of  $0.3\text{ nM}$  ( $44\text{ ng ml}^{-1}$ )), which was similar to or better than the parental IgG and an over 100-fold-better  $IC_{50}$  than that of the parental Fabs (Fig. 1g). CoV-X2 remained effective against pseudoviruses that bear escape mutations that make them resistant to the individual monoclonal antibodies<sup>14</sup> (Fig. 1g), and against a pseudovirus with RBD mutations that are found in the B.1.351 variant (which was first reported in South Africa) ( $IC_{50} = 1.3\text{ nM}$  ( $191\text{ ng ml}^{-1}$ )) (Extended Data Fig. 5). To confirm the efficacy of CoV-X2, we performed plaque-reduction neutralization assays with infectious virus. CoV-X2 efficiently neutralized wild-type SARS-CoV-2 ( $IC_{50} = 0.9\text{ nM}$ ); the B.1 variant that carries S(D614G), which was first detected in Europe ( $IC_{50} = 0.2\text{ nM}$ ); the B.1.1.7 variant, which was first observed in the UK ( $IC_{50} = 0.2\text{ nM}$ ); the P.1 variant, which was first isolated in Brazil ( $IC_{50} = 2.1\text{ nM}$ ); and the B.1.351 variant ( $IC_{50} = 12\text{ nM}$ ) (Fig. 1h). The P.1 and B.1.351 variants have almost identical mutations in the RBD; the only difference is the presence of an Asn (in B.1.351) versus Thr (in P.1) at position 417, which does not interact with CoV-X2. Nonetheless, the neutralization of B.1.351 was lower than that for P.1, which suggests that there are either some conformational differences in the RBD between the two variants or long-range effects that derive from additional mutations in the S. A similar behaviour is seen with the wild-type sequence (D614), which has a lower neutralization than S(D614G) even if no other difference is present; a plausible explanation is that a Gly at position 614 makes the CoV-X2 epitopes more accessible by favouring the ‘up’ conformation of the RBD<sup>16</sup>. We conclude that the in vitro binding and neutralizing properties of CoV-X2 make it preferable over its parental antibodies.

To assess the clinical potential of CoV-X2, we investigated its ability to protect against infection and disease in an animal model. We developed a mouse model in which human ACE2 (hACE2) is expressed by cells of the upper and lower respiratory tract after inhalation of a modified adeno-associated virus (AAV) (AAV-hACE2) (Fig. 2, Extended Data Fig. 7, Methods).

**Fig. 2: CoV-X2 protects mice transduced with AAV-hACE2 against SARS-CoV-2 disease.**



**a**, Body weight over time in mice infected with SARS-CoV-2. We transduced 13–15-week-old C57BL/6NCrl wild-type female mice with AAV-hACE2 by forced inhalation, delivering viral particles to the upper and lower respiratory tract. After >7 days, mice were infected with SARS-CoV-2 ( $1 \times 10^4$  plaque-forming units (PFU))

or received vehicle by the intranasal route. Weight was monitored daily for 8 days (SARS-CoV-2,  $n = 5$ ; control,  $n = 4$ ). Mean with s.d. **b**, Kinetics of viral burden in lungs of mice infected with SARS-CoV-2 ( $n = 5$ ) by plaque assays. Mean with s.d. Dashed line, limit of detection. **c**, Kinetics of viral RNA levels in lung samples from mice infected with SARS-CoV-2 ( $n = 5$ ) by quantitative PCR with reverse transcription (RT–qPCR). Mean with s.d. **d**, Wild-type mice were transduced with AAV-hACE2 by forced inhalation. After >7 days, mice were inoculated intraperitoneally with antibodies one day before (black arrow) or 12 h after (red arrow) being infected intranasally with SARS-CoV-2 ( $1 \times 10^4$  PFU). **e, f**, Body weight was monitored daily in mice treated 24 h before infection (**e**, C121,  $n = 9$ ; C135,  $n = 5$ ; CoV-X2,  $n = 13$ ; isotype control,  $n = 10$ ) or 12 h after infection (**f**, CoV-X2,  $n = 4$ ; control,  $n = 10$ ). Mean with s.d. is shown. **g**, Lung viral burden by plaque assay at 2 dpi (isotype control,  $n = 8$ ; CoV-X2,  $n = 5$ ; C121,  $n = 5$ ; C135,  $n = 5$ ) and 5 dpi (isotype control,  $n = 6$ ; CoV-X2,  $n = 10$ ; C121,  $n = 5$ ; C135,  $n = 5$ ). The dashed line indicates the limit of detection; mean with s.d.  $P$  value was calculated with two-tailed Student's *t*-test. **h**, Viral RNA levels in the spleen by RT–qPCR at 5 and 8 dpi (isotype control,  $n = 6$  or 8 (indicated by dots); CoV-X2,  $n = 8$  or 10 (indicated by dots)). Mean with s.d.  $P$  value, two-tailed Student's *t*-test. **i**, Photographs of lungs from infected mice (8 dpi). **j**, Histopathology and F4/80 immunohistochemistry (IHC). Haematoxylin and eosin (H&E)-stained sections of paraffin-embedded lungs from infected mice (8 dpi). Arrowheads, foamy macrophages. F4/80 IHC shows abundant macrophage infiltration in lungs of mice treated with isotype control, but not those with CoV-X2. Each image representative of two separate experiments ( $n = 3$ –5 mice per group). Scale bars, 20  $\mu\text{m}$  (inset in H&E), 50  $\mu\text{m}$  (H&E right, F4/80 IHC), 100  $\mu\text{m}$  (H&E middle), 1 mm (H&E, left).

[Source data](#)

[Full size image](#)

This approach enables the rapid production of large cohorts of mice, and has the advantage of being applicable to wild-type and mutant mouse colonies, independently of age and sex. Moreover, because AAV vectors are only weakly immunogenic and cytotoxic, the system allows for prolonged expression of hACE2<sup>17,18,19,20</sup> (Extended Data Fig. 7). SARS-CoV-2 infection of ACE2-humanized mice (hereafter, hACE2 mice) results in progressive weight loss, respiratory pathology and disease that require culling at 8 days post-infection (dpi) (Fig. 2a–c, Extended Data Fig. 7).

To evaluate the protective effect of antibodies, we treated hACE2 mice one day before SARS-CoV-2 infection (150  $\mu\text{g}$ ) (Fig. 2d,e,g,h) or 12 h after challenge (250  $\mu\text{g}$ ) (Fig. 2d,f) and monitored them over time. Upon intranasal infection with  $1 \times 10^4$  plaque-forming units of SARS-CoV-2 (SARS-CoV-2/human/Czech Republic/951/2020), mice treated with an isotype control (an antibody against the envelope protein of Zika virus)

showed weight loss starting at 3 dpi; by 8 dpi, most mice had lost approximately 25–30% of their body weight, reaching the end point for humane culling (Fig. 2e,f). We recovered infectious virus from the lungs (Fig. 2g), and detected viral RNA in the spleen (Fig. 2h) but not in the heart (data not shown). The lung pathology of these mice resembled that associated with severe COVID-19 in humans<sup>21</sup>, and was characterized by diffuse alveolar damage (50–80% of tissue area), alveolar replacement with infiltrates of immune cells and fibroblasts, thickened septa and infiltrations by activated macrophages with foamy cytoplasm (Fig. 2j). By contrast, mice treated with CoV-X2 maintained their body weight ( $P < 0.0001$  at 4–8 dpi, when compared to mice treated with isotype control) (Fig. 2e; the  $P$  values between all groups are given in Extended Data Table 1), had a reduced viral burden in lungs and viral RNA levels in spleen (Fig. 2g,h) and displayed neither macro- nor histopathological changes (diffuse alveolar damage of less than 5–10%) (Fig. 2i,j). Although infectious virus was detected at 2 dpi in all mice that were treated with C121 ( $n = 5$ ), C135 ( $n = 5$ ) or control ( $n = 8$ ), in mice that received CoV-X2 it was detected only at low levels in one mouse (out of five) (Fig. 2g). At 5 dpi, we readily recovered infectious virus from control-treated mice (5 of 6 mice), but only from 1 out of 10 mice treated with CoV-X2 (Fig. 2g). Furthermore, CoV-X2 was also protective when administered at 12 h after SARS-CoV-2 challenge (Fig. 2d,f). As none of the mice treated with CoV-X2 exhibited symptoms at any time, we conclude that CoV-X2 protects mice from the disease.

Because monotherapy with C121 or C135 monoclonal antibodies leads to virus escape in vitro<sup>14</sup>, we treated hACE2 mice with the individual antibodies and sequenced the virus. Only wild-type RBD sequences were obtained from control mice ( $n = 10$ ). In mice treated with C121, there was selection for a mutation that results in the E484D substitution (5 of 5 mice that were analysed at 8 dpi). Escape mutations at E484 generated by C121 have previously been observed in vitro<sup>14</sup>, and changes at this residue (which are also present in the B.1.351 and P.1 variants) reduce neutralization by human sera by more than tenfold<sup>22</sup>. The E484D substitution affects intermolecular hydrogen bonds at the core of the interface between C121 and the RBD, and it has previously been suggested to increase the affinity of the RBD for ACE2<sup>23</sup>. Virus containing Asp at position 484 is pathogenic, as 7 out of 9 mice treated with C121 developed disease (Fig. 2e) and only this virus was found in their lungs. By contrast (and unlike the in vitro results<sup>14</sup>), we did not observe virus evasion or pathology in mice treated with C135 ( $n = 5$ ) (Fig. 2e and data not shown). Even though we did not retrieve infectious virus (at 8 dpi) from mice treated with CoV-X2 ( $n = 13$ ) and noticed no symptoms in these mice, we detected low levels of residual viral RNA in some mice after 40 cycles of PCR amplification: in 6 of 13 mice, the virus sequence was wild type, and in 2 mice the overlapping sequencing traces were consistent with the coexistence of wild-type virus and a variant with Asp at position 484. Thus, in these 2 (out of 13) mice that contained the E484D variant, CoV-X2 remained protective even

if the mutation diluted the effective antibody concentration (presumably leaving only the C135 moiety active).

Monoclonal antibodies that target the SARS-CoV-2 S are in advanced clinical trials and show promise against COVID-19<sup>1,2</sup>. Concomitant use of multiple antibodies is preferred for increased efficacy and adding resistance against viral evasion. Indeed, the virus can escape pressure by a single antibody in vitro and—as shown here—also in mice. Moreover, RBD mutations that threaten the efficacy of single monoclonal antibodies have already been detected in virus that is circulating in minks and humans, including mutations at the C121 and C135 epitopes (Extended Data Fig. 8). One disadvantage of antibody cocktails is the requirement for a development and production capacity twice or more that required for single monoclonal antibodies, which is a considerable challenge in light of the increased demand owing to COVID-19 related vaccines and therapeutic agents (on top of the need to maintain production of biologics for other diseases)<sup>4</sup>.

Multispecific antibodies offer the advantages of cocktails in a single molecule. We have shown that the CoV-X2 bispecific antibody is more effective than the related monoclonal antibodies at inhibiting ACE2 binding; it has a sub-nanomolar IC<sub>50</sub> against a broader array of viral sequences and it protect mice from SARS-CoV-2 even when C121 (its potent parental monoclonal antibody) fails owing to viral escape. C135, the other parental monoclonal antibody, did not generate escape mutants in our mouse experiment, but readily generated them in vitro<sup>14</sup>. CoV-X2 is expected to be more resistant to viral escape as compared to monoclonal antibodies. We have shown that CoV-X2 binds and neutralizes mutants that are not recognized by its parental monoclonal antibodies, as well as variants of concern that have recently been detected in the UK<sup>12</sup>, South Africa<sup>13</sup> and Brazil<sup>24</sup>.

Unlike other multispecific antibodies<sup>25</sup>, CoV-X2 is a fully human IgG-like molecule. As such, it is suitable for further development and could be engineered to alter effector functions. For example, the Fragment crystallizable (Fc) of CoV-X2 has already been modified by introducing LALA-PG mutations<sup>26</sup> to modulate its interaction with Fc receptor and complement without affecting its antigen-binding properties. The LALA modification prevents antibody-dependent enhancement of flavivirus infection<sup>27</sup> and may be a desirable modification in the context of SARS-CoV-2, as cellular and animal experiments with coronaviruses (including SARS-CoV<sup>28,29</sup>) support the possibility of antibody-dependent enhancement. Other modifications (such as the LS modification<sup>26</sup> for increased half-life) are easily achievable. Finally, CoV-X2 is human-derived and produced in the CrossMab format, which has already shown to be safe in clinical trials<sup>30</sup>, which further supports the possibility of future development. IgG-like bispecific antibodies are therefore worth adding to the tools that we use to address SARS-CoV-2 and its future mutations.

## Methods

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.

### Protein expression and purification

To express SARS-CoV-2 S, a codon-optimized gene encoding residues 1–1208 (GenBank (<https://www.ncbi.nlm.nih.gov/genbank/>): MN908947) with proline substitutions at residues 986 and 987, a ‘GSAS’ substitution at the furin cleavage site (residues 682–685), a C-terminal T4 fibritin trimerization motif and a C-terminal 8× HisTag was synthesized and cloned into the mammalian expression vector pcDNA3.1(+) by Genscript. Codon-optimized nucleotide sequences encoding SARS-CoV-2 RBD (residues 319–591; GenBank: MN908947) with a C-terminal 8× HisTag was obtained from Genscript. hACE2 fused to the Fc region of mouse IgG at the C terminus was also synthesized and cloned into the mammalian expression vector pcDNA3.1(+). All proteins were produced by transient PEI transfection in Expi293F cells (ThermoFisher), purified from the cell supernatants 6 days after transfection by HiTrap Chelating HP (Cytiva) or HiTrap Protein A HP (Cytiva), analysed by SDS-PAGE and dynamic light scattering on a DynaPro NanoStar (Wyatt Technology, software Dynamics v.7.1.7.16). RBD and S trimer mutations were introduced in the plasmids by gene synthesis (Genscript) and purified. All proteins underwent quality control and biophysical characterization to ensure functionality, stability, lack of aggregation and batch-to-batch reproducibility.

### Monoclonal antibody production

Plasmids containing the coding sequence for the production of the monoclonal antibodies C121, C135, and C144 were prepared as previously described<sup>3</sup>. These monoclonal antibodies were produced by transient PEI transfection in Expi293F cells (ThermoFisher), purified from the cell supernatants 6 days after transfection by HiTrap Protein A HP (Cytiva) and HiLoad Superdex 200 16/60 column (Cytiva).

### Design, expression and purification of bispecific antibodies

Bispecific antibodies in the single-chain Fv format (CoV-scB1 and CoV-scB2) were designed from the sequences of the variable regions of monoclonal antibodies C144 and C135 (CoV-scB1) or C121 and C135<sup>3</sup> (CoV-scB2), following a previously described method<sup>31</sup>, N-terminal signal peptides (residues 1–19; UniProt (<https://www.uniprot.org/>): P01743) and a C-terminal 6× HisTag were added. The

constructs were synthesized and subcloned into the mammalian expression vector pcDNA3.1(+) by Genscript. The single-chain bispecific antibodies were produced by transient PEI transfection in Expi293F cells (ThermoFisher), purified from the cell supernatants 6 days after transfection by HiTrap Chelating HP (Cytiva) and HiLoad Superdex 75 16/60 column (Cytiva).

Bispecific antibodies in the CrossMAb format were designed from the sequences of the variable regions of antibodies C144 and C135 (CoV-X1) or C121 and C135<sup>3</sup> (CoV-X2). Light and heavy chain constant region sequences (UniProt P01834 and UniProt P01857) were added. The CrossMAbs were designed, as previously described<sup>10</sup>, with CH1-CL crossover in the C135 moiety. Four constructs, one for each light chain (LC(C144), LC(C121) and LC<sup>CH1-CL</sup>(C135)) and heavy chain (HC(C144), HC(C121) and HC<sup>CH1-CL</sup>(C135)) were synthesized and subcloned into the mammalian expression vector pcDNA3.1(+) by Genscript. Signal peptides were included at the N terminus of the variable sequences for expression purposes (residues 1–19; UniProt P01743 for the heavy chains and residues 1–20; UniProt P06312 for the light chains). The antibodies were produced by PEI transient transfection, plasmid ratio 1:1:1:1, in Expi293F cells (ThermoFisher), purified from the cell supernatants 6 days after transfection by HiTrap Protein A HP (Cytiva), and HiLoad Superdex 200 16/60 column (Cytiva).

All antibodies underwent quality control and biophysical characterization to ensure functionality, stability, lack of aggregation and batch-to-batch reproducibility.

## Computational modelling

CoV-scB1 and CoV-scB2 were modelled starting from the variable fragment of available experimental structures of the parental antibodies (Protein Data Bank (PDB) (<https://www.rcsb.org/>) identifiers: 7K8X for C121, 7K8Z for C135 and 7K90 for C144); the connecting linkers were manually added (Pymol); the stability and feasibility of the bispecific constructs on the S trimer were manually and visually investigated according to structural biology considerations.

CoV-X1 and CoV-X2 were assembled starting from the experimental structures of individual Fab antibodies in complex with SARS-CoV-2 (PDB 7K8X for C121, 7K8Z for C135 and 7K90 for C144). The Fc moiety was manually placed in proximity to the C terminus of each Fab heavy chain; CH1 domains were then connected to the Fc using the ALMOST toolkit<sup>32</sup>, thus obtaining the full antibody structures.

The S trimer coordinates were obtained from available experimental structures; loop regions not present in the structure were modelled using the I-TASSER suite<sup>33</sup>. PDB 6VXX served as a basis for the ‘three down’ conformation; PDB 6VYB for the ‘one RBD up’ conformation; and PDB 7A93 for the ‘two RBD up’ conformation.

Conformations not directly available (for example, trimer with ‘three RBD up’) were obtained by structural alignment and repetition of the appropriate conformation in the S monomer structures (for example, monomeric ‘RBD up’) using the PyMol software suite<sup>34</sup>

All the combinations of S trimer (three RBD down, one up, two up and three up) and antibodies (bispecific antibodies formed by combinations of C121, C135 and C144) were subjected to 400 ns or 200 ns of fully atomistic molecular dynamics simulations to obtain energetically favourable and stable conformations using GROMACS<sup>35</sup>. Calculations were performed on the CINECA-Marconi100 supercomputer.

The system was initially set up and equilibrated through standard molecular dynamics protocols: proteins were centred in a triclinic box, 0.2 nm from the edge, filled with SPCE water model and 0.15 M Na<sup>+</sup>Cl<sup>-</sup> ions using the AMBER99SB-ILDN protein force field. Energy minimization was performed to let the ions achieve a stable conformation. Temperature and pressure equilibration steps, respectively, at 310 K and 1 bar of 100 ps each were completed before performing the full molecular dynamics simulations with the above-mentioned force field. Molecular dynamics trajectory files were analysed after removal of periodic boundary conditions. The stability of each simulated complex was verified by root mean square deviation and visual analysis.

## SPR

The antibody binding properties were analysed at 25 °C on a Biacore 8K instrument (GE Healthcare) using 10 mM HEPES pH 7.4, 150 mM NaCl, 3 mM EDTA and 0.005% Tween-20 as running buffer.

SARS-CoV-2 RBD or full S, and their mutants, were immobilized on the surface of CM5 chips (Cytiva) through standard amine coupling. Increasing concentrations of antibody (3.12, 6.25, 12.5, 25 and 50 nM) were injected using a single-cycle kinetics setting; analyte responses were corrected for unspecific binding and buffer responses. Curve fitting and data analysis were performed with Biacore Insight Evaluation Software v.2.0.15.12933.

Competition experiments were performed to assess the ability of COV-X2 to bind its target with both arms. A low amount of RBD (5 nM) was immobilized on the surface of a CM5 chip through standard amine coupling. C121 or C135 antibodies were injected at a high concentration (1.5 μM) to saturate the corresponding binding sites on the RBD; CoV-X2 (200 nM) was subsequently injected. The same experimental setting was performed with a different injection order as a control: COV-X2 (1.5 μM) injected to saturate RBD binding sites and subsequent injection of C121 or C135 (200 nM).

Analysis and comparison of kinetics parameters at different RBD concentrations were performed as following. RBD was immobilized on the surface of a CM5 chip at 5, 15, 75 and 150 nM. Increasing concentrations of antibody (3.12, 6.25, 12.5, 25 and 50 nM) were injected using a single-cycle kinetics setting; analyte responses were corrected for unspecific binding and buffer responses. Curve fitting and data analysis were performed with Biacore Insight Evaluation Software v.2.0.15.12933.

## Binding inhibition of hACE2

ELISAs were used to investigate the ability of antibodies to inhibit the binding of S to hACE2. Each experiment was performed in duplicate, reporting the mean of the two replicates; error bars represent the s.d. of the measured values. We used 96-well ELISA plates, coated at 4 °C with 37 nM S, and washed and blocked with PBS + 10% FCS. Antibodies were then added either at constant saturating concentration (160 nM) or at different dilutions (starting from 340 nM and serially diluted 1 to 3) and incubated 1 h at 25 °C; after washing, hACE2–mouse Fc was added either at constant saturating concentration (160 nM) or at different dilutions (starting from 340 nM and serially diluted 1 to 2) and left 1 h at 25 °C. After further washing, bound hACE2 was detected using standard protocols with goat anti-mouse IgG coupled to alkaline phosphatase (dilution 1:500, SouthernBiotech). ELISA plates were measured with the reader software Gen5 version 1.11.5 (BioTek Instruments). Data were analysed with Microsoft Excel 2016 and GraphPad Prism version 8.4.2.

## SARS-CoV-2 pseudotyped reporter virus and pseudotyped-virus neutralization assay

A panel of plasmids expressing RBD-mutant SARS-CoV-2 S in the context of pSARS-CoV-2-S<sub>Δ19</sub> have previously been described<sup>14,36</sup>. The mutant ‘KEN’ (K417N/E484K/N501Y) was constructed in the context of a pSARS-CoV-2-S<sub>Δ19</sub> variant with a mutation in the furin cleavage site (R683G). The IC<sub>50</sub> of these pseudotypes were compared to a wild-type SARS-CoV-2 S sequence carrying R683G. Generation of SARS-CoV-2 pseudotyped HIV-1 particles was performed as previously described<sup>3,15</sup>.

The neutralization activity of the bispecific and monoclonal antibodies was measured as previously reported<sup>3,15</sup>. In brief, fourfold serially diluted antibodies were incubated with SARS-CoV-2 pseudotyped virus for 1 h at 37 °C. The mixture was subsequently incubated with 293T cells expressing ACE2 for 48 h; the cells were washed twice with PBS and lysed with Luciferase Cell Culture Lysis 5× reagent (Promega). NanoLuc activity in lysates was measured using the Nano-Glo Luciferase Assay System (Promega) with the Modulus II Microplate Reader User interface (TURNER

BioSystems). The obtained relative luminescence units (which are indicative of virus presence) were normalized to those derived from cells infected with SARS-CoV-2 pseudotyped virus in the absence of antibodies. The IC<sub>50</sub> was determined using four-parameter nonlinear regression (GraphPad Prism).

## SARS-CoV-2 virus-neutralization assay

The neutralizing activity of CoV-X2 against SARS-CoV-2 wild type, B.1 (D614G), B.1.1.7, P.1 and B.1.351 was investigated by plaque-reduction neutralization tests following a previously reported protocol<sup>37</sup>. In brief, 50 µl of antibody, starting from a concentration of 12 µg ml<sup>-1</sup> or 190 µg ml<sup>-1</sup> in a serial threefold dilution, were mixed in a flat-bottomed tissue culture microtitre plate (COSTAR) with an equal volume of 100 median tissue culture infectious dose of infectious virus that was isolated from patients with COVID-19, sequenced, titrated and incubated at 33 °C in 5% CO<sub>2</sub>. After 1 h, 3 × 10<sup>4</sup> (100 µl) Vero E6 cells (VERO C1008, Vero 76, clone 18 E6, Vero E6; ATCC CRL-1586) were added to each well. After 3 days of incubation, cells were stained with Gram's crystal violet solution (Merck) plus 5% formaldehyde 40% m/v (Carlo Erba S.p.A.) for 30 min. Microtitre plates were then washed in water. Wells were analysed to evaluate the degree of cytopathic effect compared to untreated control. Each experiment was performed in triplicate. The IC<sub>50</sub> was determined using three-parameter nonlinear regression (GraphPad Prism).

## AAV-hACE2 vector design

### Plasmid design and construction

The AAV transfer plasmid expressing hACE 2 (AAV-hACE2) was created by replacing the GFP sequence with ACE2 cDNA obtained from the HEK293 cell line (no. CRL-1573, ATCC, mycoplasma-free, population doubling lower than 13) in a pAAV-GFP plasmid (no. AAV-400, Cellbiolabs). HEK293 cells were cultured in Dulbecco's modified Eagle's medium (DMEM) (no. D5796, Sigma Aldrich) supplemented with 10% FBS (no. 10082139, Gibco) and 1% penicillin–streptomycin (no. XC-A4122100, BioSera). Cells were kept at 37 °C under 5% CO<sub>2</sub> atmosphere and 95% humidity.

To obtain the hACE2 sequence, total RNA was isolated from a confluent 10-cm<sup>2</sup> plate of HEK293 cells using RNeasy Mini Kit (Qiagen, no. 74104) according to the manufacturer's protocol and reverse-transcribed with M-MLV Reverse Transcriptase (Promega, no. M1701). The generated cDNA was used as a template for PCR amplification with a pair of primers (ACE2 forward: 5'ATGTCAAGCTTCCCTGG 3', ACE2 reverse: 5'CTAAAAGGAGGTCTAACATC 3') specific for hACE2

(NM\_021804.3) using Phusion High-Fidelity DNA Polymerase (NEB, no. M0530S). The PCR product was separated in a 1% agarose gel (SeaKem LE AGAROSE; East Port, no. 50004); the band of appropriate size (2,418 bp) was extracted using NucleoSpin Gel and PCR Clean-up (Takara, no. 740609) according to the manufacturer protocol. The extracted product was treated with DreamTaq Green DNA Polymerase (Thermo Scientific, no. EP0711) in the presence of dATP to add 3'A overhangs to the PCR product. The product was then subcloned into pGEM-T Easy Vector (Promega, no. A1360). Proper insertion of the product was assessed by HindIII (Thermo Scientific, no. ER0501) and SacI (Thermo Scientific, no. ER1132) double-digestion control. Kozak sequence and SpeI recognition site were added at the 5' and 3' end of the *ACE2* cDNA PCR product, respectively, by amplification with a specific pair of primers (hACE2\_Kozak\_Fw: 5'-CAGGGGACGATGTCAAGCTTCCTGG-3', hACE2\_Spe\_Rv: 5'-ACTAGTGATCTAAAAGGAGGT-3') using Phusion High-Fidelity DNA Polymerase (NEB, no. M0530S). The amplified product was separated in 1% agarose gel (SeaKem LE AGAROSE; East Port, no. 50004), extracted with NucleoSpin Gel and PCR Clean-up (Takara, no. 740609) according to the manufacturer's protocol. The extracted *ACE2* sequence was PCR-amplified with a pair of primers with microhomology arms (hACE2\_IF\_Fw: 5'-TTCGAACATCGATTGCAGGGGACGATGTCAAG-3', hACE2\_Spe\_IF\_Rv: 5'-GCGCTGCTCGAGGCAACTAGTGATCTAAAAGGAGGT-3') and Phusion High-Fidelity DNA Polymerase (NEB, no. M0530S), and was subsequently purified from agarose gel with NucleoSpin Gel and PCR Clean-up (Takara, no. 740609).

The GFP sequence in the pAAV-GFP control vector (no. AAV-400, Cellbiolabs) was excised by double digestion using EcoRI (Thermo Scientific, no. ER0271)/HindIII (Thermo Scientific, no. ER0501) restriction enzymes and replaced with the hACE2 sequence flanked by microhomology arms using the In-Fusion cloning system (In-Fusion HD Cloning Kit, Takara Bio Europe, no. 639648) according to the manufacturer's protocol. Proper insertion and presence of the SpeI recognition site were confirmed by double digestion using HindIII (Thermo Scientific, no. ER0501)/MluI (Thermo Scientific, no. ER0562) and SpeI (Thermo Scientific, no. ER1252), respectively. The generated AAV-hACE2 vector was sequenced (Eurofins Genomics) using the following primers: CMV\_Fw: 5'-AAATGGGCGGTAGGCCGTG-3', seq\_hACE2\_start1: 5'-TGGAGATCTGAGGTCGG-3', seq\_hACE2\_start2: 5'-TCTTCCTCCCACAGCTCCT-3', seq\_hACE2\_1: 5'-CAGTTGATTGAAGATGTGGA-3', seq\_hACE2\_2: 5'-AGAAGTGGAGGTGGATG-3', seq\_hACE2\_3: 5'-AGAACTGAAGTTGAAAAGG-3'. The produced hACE2 sequence was 100% identical to the reference hACE2 sequence (NM\_021804.3).

### Transfection of Neuro-2a cells

Neuro-2a cells (CCL-131, ATCC, mycoplasma-free, population doubling lower than 10) used for validation of AAV-hACE2 function were cultured in DMEM (no. D5796, Sigma Aldrich) supplemented with 10% FBS (no. 10082139, Gibco) and 1% penicillin–streptomycin (no. XC-A4122100, BioSera) and kept at 37 °C under 5% CO<sub>2</sub> atmosphere and 95% humidity. Three 60-mm<sup>2</sup> plates were seeded each with 4 × 10<sup>5</sup> cells and transfected the next day with the AAV-hACE2 vector using Lipofectamine 2000 Transfection Reagent (Invitrogen, no. 11668027) according to the manufacturer’s protocol for the 6-well plate transfection. Transfection with pAAV-GFP control vector (no. AAV-400, Cellbiolabs) and non-transfected HEK293 cells were used as negative controls. After 48 h, cells were collected for western blotting in RIPA buffer, supplemented with AEBSF protease inhibitor (AppliChem, no. A1421) and cComplete Mini Protease Inhibitor Cocktail (Sigma Aldrich, no. 04693124001). Expression of hACE2 in Neuro-2a cells was compared to the non-transfected HEK293 cells.

## AAV-hACE2 particle production

### AAV293 cell transfection

The AAV293 cell line (Cellbiolabs, no. AAV-100, mycoplasma-free, population doubling lower than 8) used for AAV production was cultured in DMEM (no. D5796, Sigma Aldrich) supplemented with 10% FBS (no. 10082139, Gibco), 1% penicillin–streptomycin (no. XC-A4122100, BioSera) and 1% NEAA (no. M7145, Sigma Aldrich). Cells were kept at 37 °C under 5% CO<sub>2</sub> atmosphere and 95% humidity. A day before transfection, 6.5 × 10<sup>6</sup> cells were seeded on 15-cm<sup>2</sup> cultivation plates to reach 80–90% confluence on the day of transfection. Vectors pH helper (no. 340202, Cell BioLabs), AAV Rep/Cap 2/9n (no. 112865, Addgene) and AAV-hACE2 were used for transfection in equimolar ratio. The total amount of DNA (28 µg per plate) diluted in 1 ml per plate of DMEM was mixed with linear polyethylenimine hydrochloride, M.W. 40000 PEI (no. 24765-1, Polysciences) in 1:2.7 ratio. After 20 min of incubation at room temperature, the transfection mixture was added to a cultivation plate with FBS-reduced medium (DMEM supplemented with 1% FBS) in a dropwise manner. After 5 h of incubation at 37 °C under 5% CO<sub>2</sub> atmosphere, the medium was removed and replaced with fresh complete growth medium (DMEM supplemented with 10% FBS and 1% penicillin–streptomycin).

### AAV293-hACE2 collection

Three days after transfection, both cell medium and cells were collected for AAV particle isolation. These procedures were adapted and modified from a previous publication<sup>38</sup>. The medium was collected into 50-ml centrifuge tubes and cells were

washed twice with 5 ml of PBS. Subsequently, cells were scraped in 1 ml of PBS and centrifuged at 1,000g for 10 min at 4 °C. The supernatant was added to the previously collected cell medium; the cell pellet was kept on ice during subsequent processing. The medium was centrifuged at 3,200g for 15 min at 4 °C and the supernatant was then filtered into a sterile glass bottle using 0.22-µm PES membranes. PEG-8000 (no. V3011, Promega) was added to the medium in a glass bottle in a 1:4 ratio. The mixture was stirred slowly at 4 °C for 1 h and then incubated overnight at 4 °C without stirring to allow full virus precipitation. The following day, the medium was centrifuged at 2,800g for 20 min at 4 °C and the pellet resuspended in 10 ml of PBS solution with 0.001% Pluronic F-68 non-ionic surfactant (no. 24040032, Gibco) and 200 mM NaCl (no. S5886, Sigma Aldrich) and sonicated at 50% amplitude with 4× 1-s on/15-min off pulses on ice. The cell lysate was centrifuged at 3,200g for 15 min at 4 °C.

Subsequently, 50 U ml<sup>-1</sup> of benzonase nuclease (no. E1014-25KU, Sigma Aldrich) was added to the viral suspension to degrade any residual DNA. After incubation for 1 h at 37 °C, the viral suspension was centrifuged at 2,400g for 10 min at 4 °C and the clarified supernatant was further purified.

### **AAV-hACE2 purification by iodixanol gradient ultracentrifugation**

A gradient consisting of 15% iodixanol (in 1 M NaCl, 2.7 mM MgCl<sub>2</sub>, 2 mM KCl in phosphate buffer), 25% iodixanol (in 2.7 mM MgCl<sub>2</sub>, 2 mM KCl, and 0.001% phenol red in PBS), 40% iodixanol (in 2.7 mM MgCl<sub>2</sub> and 2 mM KCl in PBS) and 0,002% phenol red (no. P3532, Sigma Aldrich) in 60% iodixanol (OptiPrep Density Gradient Medium, no. D1556, Sigma Aldrich) was prepared in QuickSeal tubes according a previous publication<sup>38</sup>. Five ml of clarified viral supernatant was carefully added on the top of the gradient and the rest of the tube was filled up with PBS.

Ultracentrifugation was carried out at 350,000g for 90 min in a pre-cooled T70i rotor at 10 °C. After ultracentrifugation, approximately 750-µl fractions were collected from the 40% iodixanol phase using an 18G needle puncturing the QuickSeal tube at the interface of the 60% and 40% iodixanol.

### **AAV-hACE2 purity validation and buffer exchange**

The purity of the collected fractions from 40% iodixanol containing AAV-hACE2 particles was assessed by SDS-PAGE. Ten µl of each collected fraction was mixed with 3.5 µl of 4× Laemmli buffer and loaded to 4–20% Mini-PROTEAN TGX Precast Protein Gel (no. 4561096, Bio-Rad). The gels were briefly washed in dH<sub>2</sub>O and stained with silver according to the manufacturer's protocol (Pierce Silver Stain Kit, no. 10096113, Thermo Scientific).

Selected AAV-hACE2 fractions were pooled and concentrated using Amicon Ultra-0.5 Centrifugal Filter Unit (molecular weight cut-off of 100 kDa). First, Amicon filter membranes were activated by incubation with 0.1% pluronic F-68, 0.01% pluronic F-68 and 200 mM NaCl, followed by 0.001% of pluronic F-68 in PBS and centrifugation at 1,900g for 5 min at 4 °C. Pooled fractions with AAV-hACE2 particles were loaded onto activated Amicon filter membranes and centrifuged at 2,600g for 5 min at 4 °C. The membranes were washed several times with 0.001% pluronic F-68 in PBS (centrifugation at 2,600g for 8 min at 4 °C) until the residual iodixanol was completely removed from the solution. To elute and concentrate the viral suspension, the membranes were covered with about 5 ml of formulation buffer and incubated for 5 min at room temperature. Amicon filters were centrifuged at 2,600g at 4 °C for approximately 1.5 min until around 0.5 ml of the formulation buffer with AAV-hACE2 particles was left. The eluate was transferred into sterile 1.5-ml tubes, quantified and stored at 4 °C for up to 2 weeks for short-term in vivo application or at –80 °C for long-term storage.

### **AAV-hACE2 titration by qPCR**

The protocol for quantification and determination of the number of genome-containing particles of AAV-hACE2 was adapted from a previous publication<sup>39</sup>, using qPCR. Purified AAV-hACE2 particles were treated with DNase I (no. EN0521, Thermo Scientific) to eliminate contaminating plasmid DNA. Serial dilutions of a AAV-hACE2 viral suspension were used as template in two separate reactions, one detecting viral inverted terminal repeat (ITR) sequences (ITR\_Fw: 5'-GGAACCCCTAGTGATGGAGTT-3', ITR\_Rv: 5'-CGGCCTCAGTGAGCGA-3') and the second hACE2 (hACE2\_Fw: 5'-CCATTGGTCTTCTGTCACCCG-3', hACE2\_Rv: 5'-AGACCATCCACCTCCACTTCTC-3'). Data analysis was performed using the LightCycler 480 Software, version 1.5. AAV concentration (the number of viral genomes in 1 µl of AAV sample) was determined by comparison to standard curves of defined concentrations of AAV-hACE2 vector. Each qPCR run was performed in triplicate; six serial dilutions of the AAV-hACE2 vector were used as positive controls and standards

### **Mouse experiments**

This study was carried out in strict accordance with the Czech national laws and guidelines on the use of experimental animals and protection of animals against cruelty (Animal Welfare Act No. 246/1992 Coll.). The protocol was approved by the Committee on the Ethics of Animal Experiments of the Institute of Parasitology, Institute of Molecular Genetics of the Czech Academy of Sciences, and of the Departmental Expert Committee for the Approval of Projects of Experiments on

Animals of the Academy of Sciences of the Czech Republic (permits 82/2020 and 101/2020).

### **Application of AAV-hACE2 viral particles to mice**

Thirteen-to-fifteen-week-old C57BL/6NCrl female mice were anaesthetized by intraperitoneal injection of ketamine and xylazine (0.1 mg per g body weight) (Biopharm) and 0.01 mg per g (Bioveta), respectively). Viral particles containing AAV-hACE2 were diluted to a final concentration of  $4 \times 10^9$  genome copies in 40 µl of PBS. This volume was applied to mice by forced inhalation. The tip of the nose was gently clipped with tweezers and the tongue gently pulled out. After the mouse started breathing through the oral cavity, 40 µl of viral suspension were applied by a 200-µl pipette tip into the oral cavity and inhaled by the mouse through the trachea into the lungs.

To access whether hACE2 was expressed in lungs, lung tissue was collected and analysed by western blot in RIPA buffer, supplemented with AEBSF protease inhibitor (AppliChem, no. A1421) and cOmplete Mini Protease Inhibitor Cocktail (Sigma Aldrich, no. 04693124001) 1, 2 and 4 weeks after application. Expression of hACE2 in AAV-hACE2 transduced mice was compared to nontreated C57BL/6NCrl mice. Histone H3 antibody (cat. no.: ab1791, Abcam, lot: GR3237685-2; 1:1,000 dilution) was used as a loading control.

### **Mouse infection**

SARS-CoV-2 (strain SARS-CoV-2/human/Czech Republic/951/2020, isolated from a clinical sample at the National Institute of Public Health, Prague), provided by J. Weber, was used for mouse infection. The virus was passaged in Vero E6 cells five times before its use in this study.

At least 7 days after application of the AAV-hACE2 virus particles, mice were infected intranasally with SARS-CoV-2 ( $1 \times 10^4$  plaque-forming units) in a total volume of 50 µl DMEM. Mice were monitored and weighted daily over an eight-day period. Treated mice were injected intraperitoneally with either 150 µg of antibodies 24 h before the infection or 250 µg 12 h after infection. Mice were killed at the indicated times after infection and their tissues collected for analysis.

### **Measurement of the viral burden**

Tissues were weighed and homogenized using Precellys 24 (Bertin Technologies) and prepared as 20% (w/v) suspension in DMEM containing 10% newborn calf serum.

The homogenates were clarified by centrifugation at 5,000g and the supernatant medium was used for plaque assay and viral RNA isolation.

Plaque assays were performed in Vero E6 cells (ATCC CRL-1586; mycoplasma-free) grown at 37 °C and 5% CO<sub>2</sub> in DMEM (no. LM-D1112/500, Biosera) supplemented with 10% FBS (no. FB-1001G/500; Biosera), and 100 U ml<sup>-1</sup> penicillin, 100 µg ml<sup>-1</sup> streptomycin (Antibiotic Antimycotic Solution; no. A5955; Sigma), and 1% l-glutamine (no. XC-T1755/100; Biosera) using a modified version of a previously published protocol<sup>40</sup>. In brief, serial dilutions of virus were prepared in 24-well tissue culture plates and cells were added to each well ( $0.6 \times 10^5$ – $1.5 \times 10^5$  cells per well). After 4 h, the suspension was overlaid with 1.5% (w/v) carboxymethylcellulose (no. C4888; Sigma) in DMEM. Following a 5-day incubation at 37 °C and 5% CO<sub>2</sub>, plates were washed with phosphate-buffered saline and the cell monolayers were stained with naphthol blue black (no. 195243; Sigma). The virus titre was expressed as plaque-forming units per ml.

RNA was isolated from tissue homogenates using the QIAamp Viral RNA mini kit (no. 52906; Qiagen) following manufacturer's instructions.

Viral RNA was quantified using EliGene COVID19 Basic a RT (no. 90077-RT-A; Elisabeth Pharmacon) according to the manufacturer's protocol. A calibration curve was constructed from four dilutions of a sample that was quantified using Quany COVID-19 kit (no. RT-25; Clonit), according to the recommendations from the manufacturer. All real-time PCR reactions were performed using a LightCycler 480 (Roche).

For sequencing, isolated RNA was used as a template for one-step RT-PCR (Qiagen OneStep RT-PCR Kit; no. 210212; Qiagen) with primers specific for SARS-CoV-2 RBD sequence (SARS-CoV-2\_seq\_FW: 5'-GCACTTGACCCTCTCAGAAC-3'; SARS-CoV-2\_seq\_RV: 5'-GACTCAGTAAGAACACCTGTGCC-3'). The reaction mixture (final volume 25 µl) contained 5 µl of QIAGEN OneStep RT-PCR buffer, 1 µl of dNTP mix, 5 µl of 5× Q-solution, 1 µl of QIAGEN OneStep RT-PCR enzyme mix, 6 µl of RNase-free water, 1.5 µl of each primer (stock concentration, 0.01 mM) and 4 µl of template RNA. The cycling conditions were as follows: reverse transcription (30 min at 50 °C), initial PCR activation (15 min at 95 °C), 3-step cycling: 40 cycles of 94 °C for 30 s, 52.6 °C for 30 s, and 72 °C for 1 min, followed by final extension (10 min at 72 °C). The PCR products were visualized in a 1.7% agarose gel in Tris-acetate-EDTA buffer. The amplified DNA was purified by using Wizard SV Gel and PCR Clean-Up System (no. A9285; Promega), according to the recommendations of the manufacturer.

The purified DNA was prepared for sequencing (Sanger method) by a commercial service (Eurofins Genomics) with the following conditions (final volume 17 µl): 15 µl of PCR product with a concentration of 5 ng µl<sup>-1</sup> and 2 µl of primer with a concentration of 10 µM. The sequencing data were analysed using BioEdit Sequence Alignment Editor, version 7.2.0.

## Histology and immunohistochemistry

Lungs were fixed in 4% PFA. Tissues were processed using a Leica ASP6025 automatic vacuum tissue processor and embedded in paraffin using a Leica EG1150 H+C embedding station. Two-µm sections were prepared using a Leica RM2255 rotary microtome and sections were stained with H&E using Leica ST5020 + CV5030 stainer and coverslipper.

To assess the presence of macrophages, a rabbit anti-mouse F4/80 monoclonal antibody (D2S9R XP<sup>R</sup> rabbit monoclonal antibody, cat. no. 70076, Cell Signaling Technology, lot 5, RRID AB\_2799771) was used at 1:800 dilution as primary antibody. The histological sections (thickness 4–5 µm) were deparaffinized in a Multistainer Leica ST5020 (Leica Biosystems). Antigens were retrieved by heating the slides in citrate buffer pH 6 (Zytomed Systems). Endogenous peroxidase was neutralized with 3% H<sub>2</sub>O<sub>2</sub>. Sections were incubated for 1 h at room temperature with a 1:800 dilution of the primary antibody. After washing they were incubated with anti-rabbit secondary antibody conjugated with HRP (Zytomed, cat. no. ZUC 032-100, lot A0880-4; no dilution). Staining of the sections was developed with a diaminobenzidine substrate kit (DAKO, Agilent) and sections were counterstained with Harris haematoxylin (Sigma Aldrich, Merck) in a Multistainer Leica.

## Biosafety statement

All work with infectious SARS-CoV-2 was performed in biosafety level 3 facilities at the Institute of Parasitology, Biology Centre of the Czech Academy of Science (Ceske Budejovice) and Veterinary Research Institute (Brno) using appropriate powered air-purifying, positive pressure respirators and protective equipment.

## Reporting summary

Further information on research design is available in the [Nature Research Reporting Summary](#) linked to this paper.

## Data availability

The data that support the findings of this study are available within the Article and its Supplementary Information. Any other data are available from the corresponding author upon reasonable request. Published data were taken from GenBank (<https://www.ncbi.nlm.nih.gov/genbank/>), UniProt (<https://www.uniprot.org/>), PDB (<https://www.rcsb.org/>) and the ViPR database (<https://www.viprbrc.org/>). Source data are provided with this paper.

# References

1. 1.

DeFrancesco, L. COVID-19 antibodies on trial. *Nat. Biotechnol.* **38**, 1242–1252 (2020).

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

2. 2.

Klasse, P. J. & Moore, J. P. Antibodies to SARS-CoV-2 and their potential for therapeutic passive immunization. *eLife* **9**, e57877 (2020).

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

3. 3.

Robbiani, D. F. et al. Convergent antibody responses to SARS-CoV-2 in convalescent individuals. *Nature* **584**, 437–442 (2020).

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

4. 4.

Ecker, D. M. & Seymour, P. in *CPhI Annual Report 2020: Postulating the Post-COVID Pharma Paradigm*, 43–49 (Informamarkets, 2020).

5. 5.

Baum, A. et al. REGN-COV2 antibodies prevent and treat SARS-CoV-2 infection in rhesus macaques and hamsters. *Science* **370**, 1110–1115 (2020).

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

6. 6.

Schäfer, A. et al. Antibody potency, effector function, and combinations in protection and therapy for SARS-CoV-2 infection in vivo. *J. Exp. Med.* **218**, e20201993 (2021).

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

7. 7.

Schlake, T. et al. mRNA: a novel avenue to antibody therapy? *Mol. Ther.* **27**, 773–784 (2019).

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

8. 8.

Tiwari, P. M. et al. Engineered mRNA-expressed antibodies prevent respiratory syncytial virus infection. *Nat. Commun.* **9**, 3999 (2018).

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

9. 9.

Barnes, C. O. et al. SARS-CoV-2 neutralizing antibody structures inform therapeutic strategies. *Nature* **588**, 682–687 (2020).

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

10. 10.

Schaefer, W. et al. Immunoglobulin domain crossover as a generic approach for the production of bispecific IgG antibodies. *Proc. Natl Acad. Sci. USA* **108**, 11187–11192 (2011).

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

11. 11.

Walls, A. C. et al. Structure, function, and antigenicity of the SARS-CoV-2 spike glycoprotein. *Cell* **181**, 281–292.e6 (2020).

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

12. 12.

Kemp, S. et al. Recurrent emergence and transmission of a SARS-CoV-2 spike deletion ΔH69/V70. Preprint at <https://doi.org/10.1101/2020.12.14.422555> (2020).

13. 13.

Tegally, H. et al. Emergence and rapid spread of a new severe acute respiratory syndrome-related coronavirus 2 (SARS-CoV-2) lineage with multiple spike mutations in South Africa. Preprint at <https://doi.org/10.1101/2020.12.21.20248640> (2020).

14. 14.

Weisblum, Y. et al. Escape from neutralizing antibodies by SARS-CoV-2 spike protein variants. *eLife* **9**, e61312 (2020).

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

15. 15.

Schmidt, F. et al. Measuring SARS-CoV-2 neutralizing antibody activity using pseudotyped and chimeric viruses. *J. Exp. Med.* **217**, e20201181 (2020).

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

16. 16.

Benton, D. J. et al. The effect of the D614G substitution on the structure of the spike glycoprotein of SARS-CoV-2. *Proc. Natl Acad. Sci. USA* **118**, e2022586118 (2021).

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

17. 17.

Han, K. et al. Lung expression of human ACE2 sensitizes the mouse to SARS-CoV-2 infection. *Am. J. Respir. Cell Mol. Biol.* **64**, 79–88 (2021).

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

18. 18.

Hassan, A. O. et al. A SARS-CoV-2 infection model in mice demonstrates protection by neutralizing antibodies. *Cell* **182**, 744–753.e4 (2020).

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

19. 19.

Sun, J. et al. Generation of a broadly useful model for COVID-19 pathogenesis, vaccination, and treatment. *Cell* **182**, 734–743.e5 (2020).

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

20. 20.

Sun, S. H. et al. A mouse model of SARS-CoV-2 infection and pathogenesis. *Cell Host Microbe* **28**, 124–133.e4 (2020).

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

21. 21.

Deshmukh, V., Motwani, R., Kumar, A., Kumari, C. & Raza, K. Histopathological observations in COVID-19: a systematic review. *J. Clin. Pathol.* **74**, 76–83 (2021).

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

22. 22.

Greaney, A. J. et al. Comprehensive mapping of mutations to the SARS-CoV-2 receptor-binding domain that affect recognition by polyclonal human serum antibodies. *Cell Host Microbe* **29**, 463–476 (2021).

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

23. 23.

Chen, J., Wang, R., Wang, M. & Wei, G. W. Mutations strengthened SARS-CoV-2 infectivity. *J. Mol. Biol.* **432**, 5212–5226 (2020).

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

24. 24.

Sabino, E. C. et al. Resurgence of COVID-19 in Manaus, Brazil, despite high seroprevalence. *Lancet* **397**, 452–455 (2021).

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

25. 25.

Dong, J. et al. Development of humanized tri-specific nanobodies with potent neutralization for SARS-CoV-2. *Sci. Rep.* **10**, 17806 (2020).

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

26. 26.

Saunders, K. O. Conceptual approaches to modulating antibody effector functions and circulation half-life. *Front. Immunol.* **10**, 1296 (2019).

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

27. 27.

Dejnirattisai, W. et al. Cross-reacting antibodies enhance dengue virus infection in humans. *Science* **328**, 745–748 (2010).

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

28. 28.

Yip, M. S. et al. Antibody-dependent infection of human macrophages by severe acute respiratory syndrome coronavirus. *Virol. J.* **11**, 82 (2014).

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

29. 29.

Yip, M. S. et al. Antibody-dependent enhancement of SARS coronavirus infection and its role in the pathogenesis of SARS. *Hong Kong Med. J.* **22** (Suppl 4), 25–31 (2016).

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

30. 30.

Klein, C. et al. Engineering therapeutic bispecific antibodies using CrossMab technology. *Methods* **154**, 21–31 (2019).

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

31. 31.

Bardelli, M. et al. A bispecific immunotweezer prevents soluble PrP oligomers and abolishes prion toxicity. *PLoS Pathog.* **14**, e1007335 (2018).

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

32. 32.

Fu, B. et al. ALMOST: an all atom molecular simulation toolkit for protein structure determination. *J. Comput. Chem.* **35**, 1101–1105 (2014).

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

33. 33.

Yang, J. et al. The I-TASSER suite: protein structure and function prediction. *Nat. Methods* **12**, 7–8 (2015).

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

34. 34.

Schrodinger. The PyMOL Molecular Graphics System, Version 1.8 (Schrodinger 2015).

35. 35.

Van Der Spoel, D. et al. GROMACS: fast, flexible, and free. *J. Comput. Chem.* **26**, 1701–1718 (2005).

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

36. 36.

Wang, Z. et al. mRNA vaccine-elicited antibodies to SARS-CoV-2 and circulating variants. *Nature* (2021).

37. 37.

Percivalle, E. et al. West Nile or Usutu virus? A three-year follow-up of humoral and cellular response in a group of asymptomatic blood donors. *Viruses* **12**, 157 (2020).

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

38. 38.

Zolotukhin, S. et al. Recombinant adeno-associated virus purification using novel methods improves infectious titer and yield. *Gene Ther.* **6**, 973–985 (1999).

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

39. 39.

Aurnhammer, C. et al. Universal real-time PCR for the detection and quantification of adeno-associated virus serotype 2-derived inverted terminal repeat sequences. *Hum. Gene Ther. Methods* **23**, 18–28 (2012).

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

40. 40.

De Madrid, A. T. & Porterfield, J. S. A simple micro-culture method for the study of group B arboviruses. *Bull. World Health Organ.* **40**, 113–121 (1969).

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

[Download references](#)

## Acknowledgements

D.F.R., L.V., Q.P.-H., F.Baldanti and L.C. have received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement no. 101003650. This work was also supported by SNF grant 31003A\_182270 (L.V.); Lions Club Monteceneri (L.V.); George Mason University Fast Grant and IRB start-up funds (D.F.R.); NIH U01 AI151698 for the United World Antiviral Research Network (UWARN) (D.F.R. and M.C.N.); NIH grant P01-AI138398-S1 (M.C.N. and P.J.B.); 2U19AI111825 (M.C.N. and D.F.R.); the Caltech Merkin Institute for Translational Research and P50 AI150464 (P.J.B.); R37-AI64003 (P.D.B.); and R01AI78788 (T.H.); P.D.B. and M.C.N. are Howard Hughes Medical Institute Investigators. The study was also supported by the Czech

Academy of Sciences and Czech Ministry of Agriculture (RVO 68378050 (R.S.) and RVO0518 (D.R.)); Czech Ministry of Education, Youth and Sports and the European Regional Development Fund (LM2018126; CZ.1.05/2.1.00/19.0395 and CZ.1.05/1.1.00/02.0109 (R.S.) and CZ.02.1.01/0.0/0.0/15\_003/0000495 (D.R.)); Czech Science Foundation (20-14325S (D.R.)); the Bulgari Women & Science Fellowship in COVID-19 Research (F. Muecksch); the EU Joint Research Centre Exploratory Research program ('NanoMicrobials'; D. Magri); and by Ricerca Finalizzata from Ministry of Health, Italy (grant no. GR-2013-02358399 (A.P.)). We are grateful for the high-performance computing resources that were provided by S. Bassini of CINECA to M. Hust, F. Bertoglio, F. Bognuda and E. Restivo. We thank V. Zatecka, V. Martinkova and L. Kutlikova for technical assistance; and V. Babak for help with statistical analyses. We are grateful to the late F. Diederich for their mentorship.

## Author information

### Author notes

1. These authors contributed equally: Raoul De Gasparo, Mattia Pedotti

## Affiliations

1. Institute for Research in Biomedicine, Università della Svizzera italiana (USI), Bellinzona, Switzerland

Raoul De Gasparo, Mattia Pedotti, Luca Simonelli, Federica Mazzola, Salvatore Di Girolamo, Filippo Bianchini, Davide F. Robbiani & Luca Varani

2. Czech Centre of Phenogenomics and Laboratory of Transgenic Models of Diseases, Institute of Molecular Genetics of the Czech Academy of Sciences, Vestec, Czech Republic

Petr Nickl, Tereza Michalcikova, Blanka Mrazkova, Natalie Polakova, Jolana Tureckova, Veronika Iatsiuk, Dagmar Zudova, Ivana Bukova, Jan Rozman, Jan Prochazka & Radislav Sedlacek

3. Laboratory of Retrovirology, The Rockefeller University, New York, NY, USA

Frauke Muecksch, Paul D. Bieniasz & Theodora Hatzioannou

4. Molecular Virology Unit, Microbiology and Virology Department, Fondazione IRCCS Policlinico San Matteo, Pavia, Italy

Irene Cassaniti, Elena Percivalle, Josè Camilla Sammartino, Federica Giardina, Stefano Gaiarsa, Antonio Piralla & Fausto Baldanti

5. Laboratory of Molecular Immunology, The Rockefeller University, New York, NY, USA

Julio C. C. Lorenzi & Michel C. Nussenzweig

6. European Commission, Joint Research Centre, Ispra, Italy

Davide Magrì, Dora Mehn, Sabrina Gioria & Luigi Calzolai

7. Veterinary Research Institute, Brno, Czech Republic

Jan Haviernik, Vaclav Honig, Andrea Fortova, Martin Palus, Petr Bednar, Petra Strakova & Daniel Ruzek

8. Institute of Parasitology, Biology Centre of the Czech Academy of Sciences, Ceske Budejovice, Czech Republic

Vaclav Honig, Martin Palus & Daniel Ruzek

9. Faculty of Science, University of South Bohemia, Ceske Budejovice, Czech Republic

Petr Bednar

10. Institute of Organic Chemistry and Biochemistry of the Czech Academy of Sciences, Prague, Czech Republic

Radim Nencka

11. Center of Biological Defense, Military Health Institute, Military Medical Agency, Techonin, Czech Republic

Oto Pavlis

12. Department of Biosciences and Nutrition, Karolinska Institutet, Huddinge, Sweden

Qiang Pan-Hammarström

13. Division of Biology and Biological Engineering, California Institute of Technology, Pasadena, CA, USA

Christopher O. Barnes & Pamela J. Bjorkman

14. Department of Clinical Surgical Diagnostic and Pediatric Sciences, Università degli Studi di, Pavia, Italy

Fausto Baldanti

15. Howard Hughes Medical Institute, The Rockefeller University, New York, NY, USA

Michel C. Nussenzweig & Paul D. Bieniasz

## Authors

1. Raoul De Gasparo

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

2. Mattia Pedotti

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

3. Luca Simonelli

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

4. Petr Nickl

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

5. Frauke Muecksch

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

6. Irene Cassaniti

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

7. Elena Percivalle

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

8. Julio C. C. Lorenzi

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

9. Federica Mazzola

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

10. Davide Magrì

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

11. Tereza Michalcikova

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

12. Jan Haviernik

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

13. Vaclav Honig

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

14. Blanka Mrazkova

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

15. Natalie Polakova

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

16. Andrea Fortova

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

17. Jolana Tureckova

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

18. Veronika Iatsiuk

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

19. Salvatore Di Girolamo

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

20. Martin Palus

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

21. Dagmar Zudova

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

22. Petr Bednar

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

23. Ivana Bukova

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

24. Filippo Bianchini

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

25. Dora Mehn

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

26. Radim Nencka

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

27. Petra Strakova

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

28. Oto Pavlis

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

29. Jan Rozman

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

30. Sabrina Gioria

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

31. Josè Camilla Sammartino

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

32. Federica Giardina

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

33. Stefano Gaiarsa

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

34. Qiang Pan-Hammarström

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

35. Christopher O. Barnes

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

36. Pamela J. Bjorkman

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

37. Luigi Calzolai

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

38. Antonio Piralla

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

39. Fausto Baldanti

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

40. Michel C. Nussenzweig

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

41. Paul D. Bieniasz

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

42. Theodora Hatzioannou

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

43. Jan Prochazka

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

44. Radislav Sedlacek

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

45. Davide F. Robbiani

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

46. Daniel Ruzek

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

47. Luca Varani

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

## Contributions

R.D.G, M. Pedotti, L.S., F. Muecksch, J.C.C.L., F. Mazzola, D. Magri, I.C., E.P., S.D.G., M. Palus, D. Mehn, S. Goria, C.O.B., F. Bianchini, J.C.S., F.G. and S. Gaiarsa designed and carried out experiments and analysed results, and produced plasmids, antibodies and viral proteins. P.N., T.M., J.H., V.H, B.M., N.P., A.F., J.T., V.I., M. Palus, D.Z., P.B., I.B., P.S. and D.R., performed mouse experiments and analysed the results. L.V., D.F.R., D.R., Q.P.-H., F. Baldanti, A.P., L.C., P.J.B., M.C.N., P.D.B. and T.H. conceived and designed study and experiments, and analysed the results. P.N., T.M., R.N., O.P., J.P., J.R. and R.S. conceived and designed the mouse model. L.V., D.F.R., D.R. and R.D.G. wrote the manuscript, with input from all co-authors.

## Corresponding authors

Correspondence to [Davide F. Robbiani](#) or [Daniel Ruzek](#) or [Luca Varani](#).

## Ethics declarations

### Competing interests

The Institute for Research in Biomedicine has filed a provisional European patent application in connection with this work, on which L.V. is inventor (PCT/EP2020/085342). The Rockefeller University has filed a provisional US patent application (US 63/021,387) on coronavirus antibodies, on which D.F.R. and M.C.N. are inventors.

## Additional information

**Peer review information** *Nature* thanks Stanley Perlman and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

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

## Extended data figures and tables

### [Extended Data Fig. 1 Neutralization of SARS-CoV-2 pseudovirus by bispecific antibodies.](#)

**a**, Schematic of the four bispecific antibodies; two in an scFv format and two in a IgG-like CrossMAb format with knob-in-hole. The parental monoclonal antibodies that form the bispecific antibodies are colour-coded: blue, C135; orange, C144; green, C121; and purple, Fc region. **b**, All four constructs neutralize SARS-CoV-2 pseudovirus in vitro at sub-nanomolar concentrations: IC<sub>50</sub> of 0.13, 0.04, 0.74 and 0.53 nM for CoV-X1, CoV-X2, CoV-scB1 and CoV-scB2, respectively. Normalized relative luminescence values (which correlate to infection) are reported versus antibody

concentration, as previously detailed<sup>15</sup>. Mean of two replicates from one representative experiment is shown.

### **Extended Data Fig. 2 CoV-X2 engages its epitopes on all RBD conformations on the S trimer.**

**a–d**, Molecular dynamics simulations of the complex between the CoV-X2 bispecific antibody and S trimers with RBD in either all down, all up or mixed up/down conformations show that CoV-X2 can engage a single RBD with both arms (**a, b**), two adjacent RBDs in the down conformation (**c**) and two RBDs in the up and down conformation (**b, d**). The complexes were subjected to up to 400 ns of fully atomistic molecular dynamics simulations to assess feasibility and stability of the bound conformations. Root-mean-squared deviations (r.m.s.d.) values are shown to indicate structural stability. S trimer is in shades of grey, RBDs in yellow (down conformation) and orange (up), the C121 and C135 moieties of CoV-X2 are in green and blue, respectively. **e**, Schematic of the computationally predicted binding modes of CoV-X2, C121 IgG and C135 IgG on the S trimer, coloured as in **a–d**. Antibodies are represented by connected circles; ACE2 is in red on the RBD if it can bind directly to a given conformation; and an arrow points to the RBD if ACE2 binding is achieved after an allowed switch to the up conformation. For example, in the three-up conformation (left), CoV-X2 can engage all the RBDs with bivalent binding, whereas C121 and C135 can achieve only monovalent binding. C135 binding does not prevent interaction with ACE2. The situation is similar in the other S conformations (two-up and one-down, two-down and one-up, and three-down); only the bispecific antibody achieves bivalent interaction and prevents ACE2 access in all conformations.

### **Extended Data Fig. 3 CoV-X2 and its parental monoclonal antibodies bind recombinant, isolated RBD and S trimer with low nanomolar affinity.**

**a**, Representative SPR traces from which the data in **b** were derived. **b**, Kinetic parameters for the binding of C121 IgG, C135 IgG and CoV-X2 to the S trimer and RBD.

**Extended Data Fig. 4 CoV-X2 binds with low-nanomolar affinity to S mutants, including some mutants that are not recognized by the parental monoclonal antibodies C121 and C135.**

**a**, SPR-derived binding affinities of CoV-X2, C121 IgG and C135 IgG to several S trimer mutants. **b**, Mutations tested in **a** are indicated by yellow spheres on the surface representation of the S trimer. The epitopes of C121 (green) and C135 (blue) are shown.

**Extended Data Fig. 5 Efficacy of CoV-X2 against the B.1.1.7 and B.1.351 variants.**

**a**, SPR traces showing binding of CoV-X2 to the RBD corresponding to wild-type SARS-CoV-2 and the B.1.1.7 and B.1.351 variants. **b**, Residues mutated in the variants are shown as red spheres on the surface representation of the S trimer. The epitopes of C121 (green) and C135 (blue) are shown. **c**, Neutralization of SARS-CoV-2 pseudoviruses expressing wild-type S, S(N501Y) and S(K417N/E484K/N501Y/R683G) (corresponding to substitutions in the RBD in the B.1.351 variant (Fig. [1h](#))) by CoV-X2. Mean of two experiments is shown.

**Extended Data Fig. 6 SPR-based avidity assays confirm that CoV-X2 can engage bivalently on a single RBD.**

**a**, CoV-X2 and monoclonal IgGs (C121 or C135) have different binding modes that are available when high or low quantities of RBD are immobilized on the surface of the SPR chip. Monoclonal antibodies have avidity effects at high RBD concentrations owing to intermolecular binding (which results in a slower  $k_d$ ), but not at low RBD concentrations as bivalent binding to a single RBD is impossible. By contrast, the bispecific antibody has avidity at both high and low concentrations, as bivalent binding to its two epitopes on a single RBD is possible.  $k_a$  is not affected by avidity. **b**, Experimental confirmation that CoV-X2 engages bivalently on a single RBD. SPR traces used to determine  $k_a$  and  $k_d$  of monoclonal

antibodies, Fab and the bispecific antibody at different concentrations of immobilized RBD (Fig. 1d) are shown. **c**, Table summarizing the SPR results plotted in Fig. 1d.  $k_a$  and  $k_d$  were normalized against the values at the highest RBD concentration.  $k_a$  and Fab  $k_d$  were unaffected by the RBD concentration, as expected.  $k_d$  became faster for the monoclonal antibodies (loss of avidity) but less so for the bispecific antibody (avidity maintained owing to simultaneous binding to two sites on a single RBD).

### Extended Data Fig. 7 Generation of the AAV-hACE2-transduced mouse model of COVID-19.

**a**, Diagram of the AAV-hACE2 plasmid and corresponding AAV vector. **b**, Western blot analysis detecting hACE2 expression in the lungs of one non-transduced control mouse (ctrl) and 12 mice transduced with two doses of AAV-hACE2 viral particles ( $5 \times 10^{10}$  or  $1 \times 10^{11}$  genome copies (GC)). Lung tissue was collected 1, 2 or 4 weeks (w) after transduction. Histone H3 was used as control for quantification (bottom). Quantitative analysis represents normalized data from membrane images (top), and was performed using ImageJ. Representative data from two independent experiments are shown. **c**, Preparation of concentrated AAV-hACE2. AAV-hACE2 plasmid was co-transfected with pHelper and AAV Rep/Cap 2/9n vectors into 293AAV cells (Methods). To increase viral titres, viral particles from both cell lysate and PEG-precipitated growth medium were ultracentrifuged in a discontinuous iodixanol gradient. The silver-stained SDS-PAGE gel shows 14 consecutive fractions: fractions 1–9 represent enriched AAV fractions used for experiments, and fractions 10–14 are contaminated with proteinaceous cell debris. Iodixanol was chosen as a density gradient medium owing to its low toxicity in vivo and its easy removal by ultrafiltration. M, protein marker; \*AAV capsid proteins VP1, VP2, and VP3. Representative data from two independent experiments are shown. **d**, The amount of AAV particles was estimated by RT-qPCR. The number of genome copies expressed as log was calculated from a standard curve. From one 15-cm<sup>2</sup> dish, 75 µl with  $2.0 \times 10^{12}$  genome copies per ml were prepared, which is sufficient for hACE2 humanization of 37 mice. **e**, Kinetic of lung histopathology in SARS-CoV-2-infected ACE2-humanized mice. H&E-stained sections showed inflammatory infiltrates composed of

lymphocytes, macrophages, neutrophils and fibroblasts replacing the alveoli. The size of the affected areas increased over time (area of diffuse alveolar damage: control <5–10%, 2 dpi <10–30%, 5 dpi 20–80% and 8 dpi 50–90%). Alveolar septa were thickened in areas that were close to infiltrates. In samples collected at 5 and 8 dpi, an increased number of activated macrophages with foamy cytoplasm (black arrowheads) was seen. AAV-hACE2-transduced, SARS-CoV-2-uninfected mice were used as control and showed no noticeable pathology. Each image is representative of two separate experiments ( $n = 3$  to 5 mice per group).

### **Extended Data Fig. 8 Natural SARS-CoV-2 variants in the C121 and C135 epitopes.**

**a, b**, Summary of naturally occurring mutations in the C121 (**a**) or C135 (**b**) epitopes that have been reported in circulating SARS-CoV-2 (as of 1 January 2021). The location of the mutated residues is shown in red on the RBD structure. The C121 and C135 variable regions are in green and blue, respectively (PDB 7K8X and 7K8Z, respectively). All the variants were taken from the ViPR database (<https://www.viprbrc.org/>).

### **Extended Data Table 1 Summary of the *P* values for the mouse protection experiment**

[Full size table](#)

## **Supplementary information**

### **Supplementary Figure 1**

Generation of the new AAV-hACE2-transduced mouse model for COVID-19; western blot analysis detecting hACE2 expression in the lungs of one non-transduced control mouse (Ctrl) and 12 mice transduced with two different doses of AAV-hACE2 viral particles ( $5 \times 10^{10}$  or  $1 \times 10^{11}$  genome copies, GC). Lung tissue was collected 1, 2, or 4 weeks (w) post transduction. Histone H3 was used as control for quantification (bottom). Quantitative analysis represents normalized data from membrane images (top). Western blot and quantitative analysis are derived from the same

experiment on a single gel and blot on the same membrane. The membrane was cut horizontally at the level of 25 kDa, since no ACE2 products below that are expected; the top part was stained with anti-ACE2 antibody, the lower with anti-Histone H3 antibody. Staining for molecular weight marker (right) was performed on the same gel and reported on the left for convenience.

## [Reporting Summary](#)

## **Source data**

### [Source Data Fig. 2](#)

## **Rights and permissions**

### [Reprints and Permissions](#)

## **About this article**



## **Cite this article**

De Gasparo, R., Pedotti, M., Simonelli, L. *et al.* Bispecific IgG neutralizes SARS-CoV-2 variants and prevents escape in mice. *Nature* **593**, 424–428 (2021). <https://doi.org/10.1038/s41586-021-03461-y>

### [Download citation](#)

- Received: 07 January 2021
- Accepted: 16 March 2021

- Published: 25 March 2021
- Issue Date: 20 May 2021
- DOI: <https://doi.org/10.1038/s41586-021-03461-y>

## Further reading

- **Unveiling the Potential Role of Nanozymes in Combating the COVID-19 Outbreak**
  - Jafar Ali
  - , Saira Naveed Elahi
  - , Asghar Ali
  - , Hassan Waseem
  - , Rameesha Abid
  - & Mohamed M. Mohamed

*Nanomaterials* (2021)
- **COVID-19 one year into the pandemic: from genetics and genomics to therapy, vaccination, and policy**
  - Giuseppe Novelli
  - , Michela Biancolella
  - , Ruty Mehrian-Shai
  - , Vito Luigi Colona
  - , Anderson F. Brito
  - , Nathan D. Grubaugh
  - , Vasilis Vasilious
  - , Lucio Luzzatto
  - & Juergen K. V. Reichardt

*Human Genomics* (2021)

## Comments

By submitting a comment you agree to abide by our [Terms](#) and [Community Guidelines](#). If you find something abusive or that does not comply with our terms or guidelines please flag it as inappropriate.

[Download PDF](#)

Advertisement

---

This article was downloaded by **calibre** from <https://www.nature.com/articles/s41586-021-03461-y>.

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

- Article
- [Published: 19 May 2021](#)

# In vivo CRISPR base editing of *PCSK9* durably lowers cholesterol in primates

- [Kiran Musunuru](#) ORCID: [orcid.org/0000-0003-3298-0368<sup>1,2,3</sup>](https://orcid.org/0000-0003-3298-0368),
- [Alexandra C. Chadwick<sup>4</sup>](#),
- [Taiji Mizoguchi<sup>4</sup>](#),
- [Sara P. Garcia](#) ORCID: [orcid.org/0000-0003-3312-6454<sup>4</sup>](https://orcid.org/0000-0003-3312-6454),
- [Jamie E. DeNizio<sup>4</sup>](#),
- [Caroline W. Reiss](#) ORCID: [orcid.org/0000-0002-6385-1879<sup>4</sup>](https://orcid.org/0000-0002-6385-1879),
- [Kui Wang<sup>4</sup>](#),
- [Sowmya Iyer<sup>4</sup>](#),
- [Chaitali Dutta<sup>4</sup>](#),
- [Victoria Clendaniel](#) ORCID: [orcid.org/0000-0002-2969-4443<sup>4</sup>](https://orcid.org/0000-0002-2969-4443),
- [Michael Amaonye](#) ORCID: [orcid.org/0000-0001-6683-2545<sup>4</sup>](https://orcid.org/0000-0001-6683-2545),
- [Aaron Beach<sup>4</sup>](#),
- [Kathleen Berth<sup>4</sup>](#),
- [Souvik Biswas<sup>4</sup>](#),
- [Maurine C. Braun<sup>4</sup>](#),
- [Huei-Mei Chen<sup>4</sup>](#),
- [Thomas V. Colace<sup>4</sup>](#),
- [John D. Ganey<sup>4</sup>](#),
- [Soumyashree A. Gangopadhyay<sup>4</sup>](#),
- [Ryan Garrity](#) ORCID: [orcid.org/0000-0002-5015-4505<sup>4</sup>](https://orcid.org/0000-0002-5015-4505),
- [Lisa N. Kasiewicz<sup>4</sup>](#),
- [Jennifer Lavoie<sup>4</sup>](#),
- [James A. Madsen<sup>4</sup>](#),

- [Yuri Matsumoto<sup>4</sup>](#),
- [Anne Marie Mazzola<sup>4</sup>](#),
- [Yusuf S. Nasrullah<sup>4</sup>](#),
- [Joseph Nneji<sup>4</sup>](#),
- [Huilan Ren<sup>4</sup>](#),
- [Athul Sanjeev ORCID: orcid.org/0000-0003-4089-5315<sup>4</sup>](#),
- [Madeleine Shay<sup>4</sup>](#),
- [Mary R. Stahley<sup>4</sup>](#),
- [Steven H. Y. Fan<sup>5</sup>](#),
- [Ying K. Tam<sup>5</sup>](#),
- [Nicole M. Gaudelli ORCID: orcid.org/0000-0001-8608-4855<sup>6</sup>](#),
- [Giuseppe Ciaramella<sup>6</sup>](#),
- [Leslie E. Stolz<sup>4</sup>](#),
- [Padma Malyala<sup>4</sup>](#),
- [Christopher J. Cheng<sup>4</sup>](#),
- [Kallanthottathil G. Rajeev<sup>4</sup>](#),
- [Ellen Rohde<sup>4</sup>](#),
- [Andrew M. Bellinger<sup>4</sup>](#) &
- [Sekar Kathiresan ORCID: orcid.org/0000-0002-3711-7101<sup>4</sup>](#)

[Nature](#) volume **593**, pages 429–434 (2021) [Cite this article](#)

- 6537 Accesses
- 794 Altmetric
- [Metrics details](#)

## Subjects

- [Drug delivery](#)
- [Dyslipidaemias](#)
- [Gene therapy](#)
- [Genetic engineering](#)
- [Nucleic-acid therapeutics](#)

# Abstract

Gene-editing technologies, which include the CRISPR–Cas nucleases<sup>1,2,3</sup> and CRISPR base editors<sup>4,5</sup>, have the potential to permanently modify disease-causing genes in patients<sup>6</sup>. The demonstration of durable editing in target organs of nonhuman primates is a key step before in vivo administration of gene editors to patients in clinical trials. Here we demonstrate that CRISPR base editors that are delivered in vivo using lipid nanoparticles can efficiently and precisely modify disease-related genes in living cynomolgus monkeys (*Macaca fascicularis*). We observed a near-complete knockdown of *PCSK9* in the liver after a single infusion of lipid nanoparticles, with concomitant reductions in blood levels of PCSK9 and low-density lipoprotein cholesterol of approximately 90% and about 60%, respectively; all of these changes remained stable for at least 8 months after a single-dose treatment. In addition to supporting a ‘once-and-done’ approach to the reduction of low-density lipoprotein cholesterol and the treatment of atherosclerotic cardiovascular disease (the leading cause of death worldwide<sup>7</sup>), our results provide a proof-of-concept for how CRISPR base editors can be productively applied to make precise single-nucleotide changes in therapeutic target genes in the liver, and potentially in other organs.

## Access options

Subscribe to Journal

Get full journal access for 1 year

\$199.00

only \$3.90 per issue

[Subscribe](#)

All prices are NET prices.

VAT will be added later in the checkout.

Tax calculation will be finalised during checkout.

Rent or Buy article

Get time limited or full article access on ReadCube.

from \$8.99

[Rent or Buy](#)

All prices are NET prices.

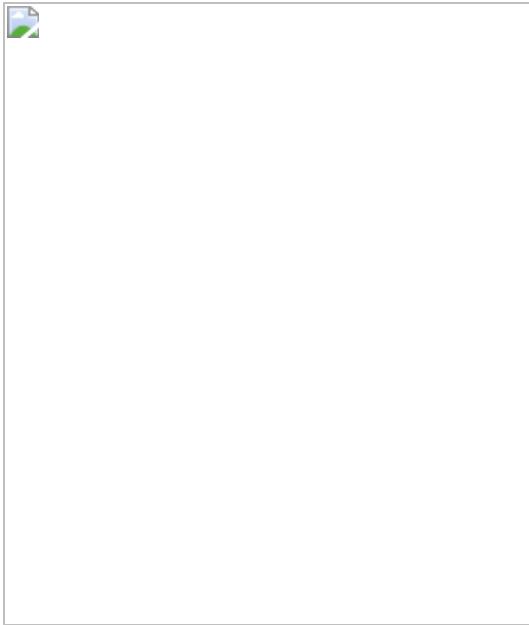
### **Additional access options:**

- [Log in](#)
- [Access through your institution](#)
- [Learn about institutional subscriptions](#)

**Fig. 1: Adenine base editing of *PCSK9* in primary human hepatocytes, primary cynomolgus monkey hepatocytes and mice.**



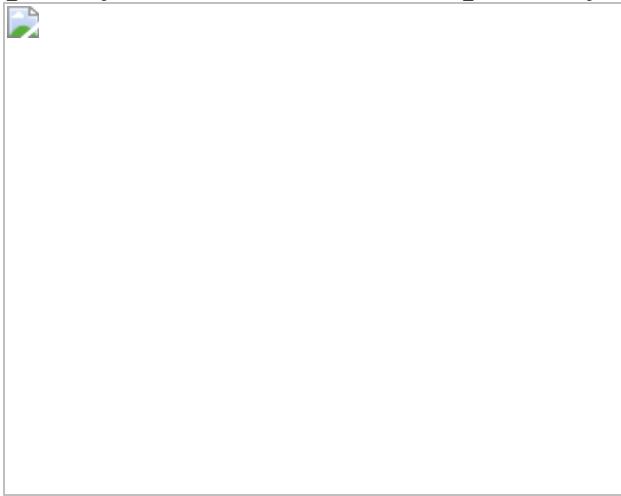
**Fig. 2: Short-term effects of adenine base editing of *PCSK9* in cynomolgus monkeys.**



**Fig. 3: Long-term effects of adenine base editing of *PCSK9* in cynomolgus monkeys.**



**Fig. 4: Assessment of off-target editing in primary cynomolgus monkey hepatocytes and liver, and in primary human hepatocytes.**



## Data availability

DNA and RNA sequencing data that support the findings of this study have been deposited in the NCBI Sequence Read Archive with the accession code [PRJNA716270](#). All other data supporting the findings of this study

(Figs. 1–4, Extended Data Figs. 1–9) are available within the Article and its Supplementary Information. The GRCh38 reference human genome ([ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000/001/405/GCA\\_000001405.15\\_GRCh38/seqs\\_for\\_alignment\\_pipelines.ucsc\\_ids/GCA\\_000001405.15\\_GRCh38\\_no\\_alt\\_analysis\\_set.fna.gz](ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000/001/405/GCA_000001405.15_GRCh38/seqs_for_alignment_pipelines.ucsc_ids/GCA_000001405.15_GRCh38_no_alt_analysis_set.fna.gz), [ftp://ftp.ensembl.org/pub/release-98/fasta/homo\\_sapiens/dna/Homo\\_sapiens.GRCh38.dna.chromosome.{1-22,X,Y,MT}.fa](ftp://ftp.ensembl.org/pub/release-98/fasta/homo_sapiens/dna/Homo_sapiens.GRCh38.dna.chromosome.{1-22,X,Y,MT}.fa) and [ftp://ftp.ensembl.org/pub/release-98/fasta/homo\\_sapiens/dna/Homo\\_sapiens.GRCh38.dna.nonchromosomal.fa](ftp://ftp.ensembl.org/pub/release-98/fasta/homo_sapiens/dna/Homo_sapiens.GRCh38.dna.nonchromosomal.fa)) and Gencode v.34 ([ftp://ftp.ebi.ac.uk/pub/databases/gencode/Gencode\\_human/release\\_34/gencode.v34.primary\\_assembly.annotation.gtf.gz](ftp://ftp.ebi.ac.uk/pub/databases/gencode/Gencode_human/release_34/gencode.v34.primary_assembly.annotation.gtf.gz)) and Ensembl v.98 ([ftp://ftp.ensembl.org/pub/release-98/gtf/homo\\_sapiens/Homo\\_sapiens.GRCh38.98.gtf.gz](ftp://ftp.ensembl.org/pub/release-98/gtf/homo_sapiens/Homo_sapiens.GRCh38.98.gtf.gz)) annotations were used. The macFas5 cynomolgus monkey reference genome ([ftp://ftp.ensembl.org/pub/release-98/fasta/macaca\\_fascicularis/dna/Macaca\\_fascicularis.Macaca\\_fascicularis\\_5.0.dna.chromosome.{1-20,X,MT}.fa.gz](ftp://ftp.ensembl.org/pub/release-98/fasta/macaca_fascicularis/dna/Macaca_fascicularis.Macaca_fascicularis_5.0.dna.chromosome.{1-20,X,MT}.fa.gz) and [ftp://ftp.ensembl.org/pub/release-98/fasta/macaca\\_fascicularis/dna/Macaca\\_fascicularis.Macaca\\_fascicularis\\_5.0.dna.nonchromosomal.fa.gz](ftp://ftp.ensembl.org/pub/release-98/fasta/macaca_fascicularis/dna/Macaca_fascicularis.Macaca_fascicularis_5.0.dna.nonchromosomal.fa.gz)) was used. Source data are provided with this paper.

## Code availability

Custom codes used to analyse Digenome-seq data are provided in the Supplementary Information (files named `digenome_step1.sh` and `digenome_step2.R`), as are instructions (file named `README.txt`).

## References

1. 1.

Jinek, M. et al. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **337**, 816–821 (2012).

[ADS](#) [CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

2. 2.

Zetsche, B. et al. Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system. *Cell* **163**, 759–771 (2015).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

3. 3.

Strecker, J. et al. Engineering of CRISPR–Cas12b for human genome editing. *Nat. Commun.* **10**, 212 (2019).

[ADS](#) [CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

4. 4.

Komor, A. C., Kim, Y. B., Packer, M. S., Zuris, J. A. & Liu, D. R. Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature* **533**, 420–424 (2016).

[ADS](#) [CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

5. 5.

Gaudelli, N. M. et al. Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage. *Nature* **551**, 464–471 (2017).

[ADS](#) [CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

6. 6.

Doudna, J. A. The promise and challenge of therapeutic genome editing. *Nature* **578**, 229–236 (2020).

[ADS](#) [CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

7. 7.

GBD 2017 Causes of Death Collaborators. Global, regional, and national age-sex-specific mortality for 282 causes of death in 195 countries and territories, 1980-2017: a systematic analysis for the Global Burden of Disease Study 2017. *Lancet* **392**, 1736–1788 (2018).

[Google Scholar](#)

8. 8.

Anzalone, A. V. et al. Search-and-replace genome editing without double-strand breaks or donor DNA. *Nature* **576**, 149–157 (2019).

[ADS](#) [CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

9. 9.

Yin, H. et al. Genome editing with Cas9 in adult mice corrects a disease mutation and phenotype. *Nat. Biotechnol.* **32**, 551–553 (2014).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

10. 10.

Ding, Q. et al. Permanent alteration of *PCSK9* with in vivo CRISPR-Cas9 genome editing. *Circ. Res.* **115**, 488–492 (2014).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

11. 11.

Yin, H. et al. Structure-guided chemical modification of guide RNA enables potent non-viral in vivo genome editing. *Nat. Biotechnol.* **35**, 1179–1187 (2017).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

12. 12.

Finn, J. D. et al. A single administration of CRISPR/Cas9 lipid nanoparticles achieves robust and persistent in vivo genome editing. *Cell Rep.* **22**, 2227–2235 (2018).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

13. 13.

Chadwick, A. C., Wang, X. & Musunuru, K. In vivo base editing of *PCSK9* (proprotein convertase subtilisin/kexin type 9) as a therapeutic alternative to genome editing. *Arterioscler. Thromb. Vasc. Biol.* **37**, 1741–1747 (2017).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

14. 14.

Ryu, S. M. et al. Adenine base editing in mouse embryos and an adult mouse model of Duchenne muscular dystrophy. *Nat. Biotechnol.* **36**, 536–539 (2018).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

15. 15.

Rossidis, A. C. et al. In utero CRISPR-mediated therapeutic editing of metabolic genes. *Nat. Med.* **24**, 1513–1518 (2018).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

16. 16.

Abifadel, M. et al. Mutations in *PCSK9* cause autosomal dominant hypercholesterolemia. *Nat. Genet.* **34**, 154–156 (2003).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

17. 17.

Cohen, J. C., Boerwinkle, E., Mosley, T. H., Jr & Hobbs, H. H. Sequence variations in *PCSK9*, low LDL, and protection against coronary heart disease. *N. Engl. J. Med.* **354**, 1264–1272 (2006).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

18. 18.

Rao, A. S. et al. Large-scale phenome-wide association study of *PCSK9* variants demonstrates protection against ischemic stroke. *Circ. Genom. Precis. Med.* **11**, e002162 (2018).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

19. 19.

Zhao, Z. et al. Molecular characterization of loss-of-function mutations in *PCSK9* and identification of a compound heterozygote. *Am. J. Hum. Genet.* **79**, 514–523 (2006).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

20. 20.

Hooper, A. J., Marais, A. D., Tanyanyiwa, D. M. & Burnett, J. R. The C679X mutation in *PCSK9* is present and lowers blood cholesterol in a southern African population. *Atherosclerosis* **193**, 445–448 (2007).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

21. 21.

Ray, K. K. et al. Inclisiran in patients at high cardiovascular risk with elevated LDL cholesterol. *N. Engl. J. Med.* **376**, 1430–1440 (2017).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

22. 22.

Brandts, J. & Ray, K. K. Low density lipoprotein cholesterol-lowering strategies and population health: time to move to a cumulative exposure model. *Circulation* **141**, 873–876 (2020).

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

23. 23.

Choudhry, N. K. et al. Full coverage for preventive medications after myocardial infarction. *N. Engl. J. Med.* **365**, 2088–2097 (2011).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

24. 24.

Rodriguez, F. et al. Association of statin adherence with mortality in patients with atherosclerotic cardiovascular disease. *JAMA Cardiol.* **4**, 206–213 (2019).

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

25. 25.

Hines, D. M., Rane, P., Patel, J., Harrison, D. J. & Wade, R. L. Treatment patterns and patient characteristics among early initiators of PCSK9 inhibitors. *Vasc. Health Risk Manag.* **14**, 409–418 (2018).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

26. 26.

Zafrir, B., Egbaria, A., Stein, N., Elis, A. & Saliba, W. PCSK9 inhibition in clinical practice: treatment patterns and attainment of lipid goals in a large health maintenance organization. *J. Clin. Lipidol.* **15**, 202–211.e2 (2021).

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

27. 27.

Gaudelli, N. M. et al. Directed evolution of adenine base editors with increased activity and therapeutic application. *Nat. Biotechnol.* **38**, 892–900 (2020).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

28. 28.

Hendel, A. et al. Chemically modified guide RNAs enhance CRISPR–Cas genome editing in human primary cells. *Nat. Biotechnol.* **33**, 985–989 (2015).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

29. 29.

Conway, A. et al. Non-viral delivery of zinc finger nuclease mRNA enables highly efficient *in vivo* genome editing of multiple therapeutic gene targets. *Mol. Ther.* **27**, 866–877 (2019).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

30. 30.

Villiger, L. et al. *In vivo* cytidine base editing of hepatocytes without detectable off-target mutations in RNA and DNA. *Nat. Biomed. Eng.* **5**, 179–189 (2021).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

31. 31.

Petri, K. et al. Global-scale CRISPR gene editor specificity profiling by ONE-seq identifies population-specific, variant off-target effects. Preprint at <https://doi.org/10.1101/2021.04.05.438458> (2021).

32. 32.

Liang, P. et al. Genome-wide profiling of adenine base editor specificity by EndoV-seq. *Nat. Commun.* **10**, 67 (2019).

[ADS](#) [CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

33. 33.

Kim, D., Kim, D. E., Lee, G., Cho, S. I. & Kim, J. S. Genome-wide target specificity of CRISPR RNA-guided adenine base editors. *Nat. Biotechnol.* **37**, 430–435 (2019).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

34. 34.

Grünewald, J. et al. Transcriptome-wide off-target RNA editing induced by CRISPR-guided DNA base editors. *Nature* **569**, 433–437 (2019).

[ADS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

35. 35.

Zhou, C. et al. Off-target RNA mutation induced by DNA base editing and its elimination by mutagenesis. *Nature* **571**, 275–278 (2019).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

36. 36.

Wang, L. et al. Long-term stable reduction of low-density lipoprotein in nonhuman primates following in vivo genome editing of PCSK9. *Mol. Ther.* <https://doi.org/10.1016/j.ymthe.2021.02.020> (2021).

37. 37.

Villiger, L. et al. Treatment of a metabolic liver disease by in vivo genome base editing in adult mice. *Nat. Med.* **24**, 1519–1525 (2018).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

38. 38.

Song, C. Q. et al. Adenine base editing in an adult mouse model of tyrosinaemia. *Nat. Biomed. Eng.* **4**, 125–130 (2020).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

39. 39.

Koblan, L. W. et al. In vivo base editing rescues Hutchinson–Gilford progeria syndrome in mice. *Nature* **589**, 608–614 (2021).

[ADS](#) [CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

40. 40.

Clement, K. et al. CRISPResso2 provides accurate and rapid genome editing sequence analysis. *Nat. Biotechnol.* **37**, 224–226 (2019).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

41. 41.

Magoč, T. & Salzberg, S. L. FLASH: fast length adjustment of short reads to improve genome assemblies. *Bioinformatics* **27**, 2957–2963 (2011).

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

42. 42.

Concordet, J. P. & Haeussler, M. CRISPOR: intuitive guide selection for CRISPR/Cas9 genome editing experiments and screens. *Nucleic Acids Res.* **46**, W242–W245 (2018).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

43. 43.

Van der Auwera, G. A. et al. From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. *Curr. Protoc. Bioinformatics* **43**, 11.10.1–11.10.33 (2013).

[Google Scholar](#)

44. 44.

Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).

[CAS](#) [Google Scholar](#)

45. 45.

Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).

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

46. 46.

Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).

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

47. 47.

Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

48. 48.

Li, H. et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).

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

49. 49.

Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

50. 50.

Bae, S., Park, J. & Kim, J. S. Cas-OFFinder: a fast and versatile algorithm that searches for potential off-target sites of Cas9 RNA-guided endonucleases. *Bioinformatics* **30**, 1473–1475 (2014).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

[Download references](#)

## Acknowledgements

Conception and design of the work was performed and supported by Verve Therapeutics. Acquisition, analysis and interpretation of the data was performed and supported by Verve Therapeutics, with some aspects being performed on behalf or at the direction of Verve Therapeutics. Acuitas Therapeutics supported the work by providing Verve Therapeutics with LNP reagents and manufacturing the LNP formulations. Beam Therapeutics supported the work by developing the ABE8.8-m protein sequence. We are grateful to J. K. Joung for critical reading of the manuscript.

## Author information

### Affiliations

1. Cardiovascular Institute, Perelman School of Medicine at the University of Pennsylvania, Philadelphia, PA, USA

Kiran Musunuru

2. Division of Cardiovascular Medicine, Department of Medicine, Perelman School of Medicine at the University of Pennsylvania, Philadelphia, PA, USA

Kiran Musunuru

3. Department of Genetics, Perelman School of Medicine at the University of Pennsylvania, Philadelphia, PA, USA

Kiran Musunuru

4. Verve Therapeutics, Cambridge, MA, USA

Alexandra C. Chadwick, Taiji Mizoguchi, Sara P. Garcia, Jamie E. DeNizio, Caroline W. Reiss, Kui Wang, Sowmya Iyer, Chaitali Dutta, Victoria Clendaniel, Michael Amaonye, Aaron Beach, Kathleen Berth, Souvik Biswas, Maurine C. Braun, Huei-Mei Chen, Thomas V. Colace, John D. Ganey, Soumyashree A. Gangopadhyay, Ryan Garrity, Lisa N. Kasiewicz, Jennifer Lavoie, James A. Madsen, Yuri Matsumoto, Anne Marie Mazzola, Yusuf S. Nasrullah, Joseph Nneji, Huilan Ren, Athul Sanjeev, Madeleine Shay, Mary R. Stahley, Leslie E. Stoltz, Padma Malyala, Christopher J. Cheng, Kallanthottathil G. Rajeev, Ellen Rohde, Andrew M. Bellinger & Sekar Kathiresan

5. Acuitas Therapeutics, Vancouver, British Columbia, Canada

Steven H. Y. Fan & Ying K. Tam

6. Beam Therapeutics, Cambridge, MA, USA

Nicole M. Gaudelli & Giuseppe Ciaramella

Authors

1. Kiran Musunuru

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

2. Alexandra C. Chadwick

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

3. Taiji Mizoguchi

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

4. Sara P. Garcia

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

5. Jamie E. DeNizio

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

6. Caroline W. Reiss

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

7. Kui Wang

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

8. Sowmya Iyer

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

9. Chaitali Dutta

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

10. Victoria Clendaniel

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

11. Michael Amaonye

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

12. Aaron Beach

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

13. Kathleen Berth

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

14. Souvik Biswas

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

15. Maurine C. Braun

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

16. Huei-Mei Chen

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

17. Thomas V. Colace

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

18. John D. Ganey

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

19. Soumyashree A. Gangopadhyay

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

20. Ryan Garrity

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

21. Lissa N. Kasiewicz

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

22. Jennifer Lavoie

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

23. James A. Madsen

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

24. Yuri Matsumoto

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

25. Anne Marie Mazzola

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

26. Yusuf S. Nasrullah

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

27. Joseph Nneji

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

28. Huilan Ren

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

29. Athul Sanjeev

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

30. Madeleine Shay

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

31. Mary R. Stahley

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

32. Steven H. Y. Fan

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

33. Ying K. Tam

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

34. Nicole M. Gaudelli

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

35. Giuseppe Ciaramella

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

36. Leslie E. Stoltz

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

37. Padma Malyala

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

38. Christopher J. Cheng

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

39. Kallanthottathil G. Rajeev

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

40. Ellen Rohde

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

41. Andrew M. Bellinger

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

42. Sekar Kathiresan

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

## Contributions

A.C.C. and K.G.R. conceived, designed and directed the gRNA design and screening. A.C.C. conceived, designed and directed the cell-based studies and off-target analyses. E.R. conceived, designed and directed the mouse and cynomolgus studies with design input from P.M. C.J.C. and C.W.R. conceived, designed and directed mRNA modification and codon optimization. C.J.C, K.B., C.W.R, A.S. and K.W. conceived, designed and/or directed mRNA processing optimization. K.M., A.C.C., T.M., J.E.D., C.W.R., K.W., C.D., V.C., M.A., A.B., K.B., S.B., M.C.B., H.-M.C., T.V.C., J.D.G., S.A.G., R.G., L.N.K., J.L., J.A.M., Y.M., A.M.M., Y.S.N., J.N., H.R., A.S., M.S., M.R.S., L.E.S., K.G.R., P.M., C.J.C. and E.R. contributed to wet laboratory experiments. K.M., S.P.G. and S.I. contributed to bioinformatic analyses. S.H.Y.F. and Y.K.T. contributed to the formulation and manufacture of LNPs. N.M.G. and G.C. contributed to the development of base-editing technology and specifically the ABE8.8-m protein sequence. K.M. wrote the manuscript, and all authors contributed to the editing of the manuscript. A.M.B. supervised the work with oversight by S.K. and advisory input from K.M.

## Corresponding author

Correspondence to [Sekar Kathiresan](#).

## Ethics declarations

## Competing interests

K.M. is an advisor to and holds equity in Verve Therapeutics and Variant Bio. A.M.B. is an employee of Verve Therapeutics and holds equity in Verve Therapeutics, Lyndra Therapeutics, Corner Therapeutics and Cocoon Biotech. S.K. is an employee of Verve Therapeutics, holds equity in Verve Therapeutics and Maze Therapeutics, and has served as a consultant for Acceleron, Eli Lilly, Novartis, Merck, Novo Nordisk, Novo Ventures, Ionis, Alnylam, Aegerion, Haug Partners, Noble Insights, Leerink Partners, Bayer Healthcare, Illumina, Color Genomics, MedGenome, Quest and Medscape. S.H.Y.F. and Y.K.T. are employees of and hold equity in Acuitas Therapeutics. N.M.G. and G.C. are employees of and hold equity in Beam Therapeutics. All other authors are employees of and hold equity in Verve Therapeutics. Verve Therapeutics has filed for patent protection related to various aspects of therapeutic base editing of *PCSK9*, with A.C.C., C.J.C., C.W.R., K.G.R. and E.R. as the inventors.

## Additional information

**Peer review information** *Nature* thanks Kathryn Moore, Alan Tall, Fyodor Urnov and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

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

## Extended data figures and tables

### [Extended Data Fig. 1 Overview of base-editing approach.](#)

**a**, Schematic of adenine base editing. **b**, Schematic showing potential splicing outcomes with disruption of splice donor or splice acceptor sequences. Other outcomes are possible, such as inclusion of part of the intron in the splicing product. **c**, Schematic with Sanger sequencing chromatogram, demonstrating editing of adenine base in the antisense strand at the splice donor at the end of *PCSK9* exon 1 (PCR amplification from the genomic DNA of the cells transfected with a dose of 2,500 ng ml<sup>-1</sup> in Fig. 1b), portraying how splice-site disruption results in an in-frame

stop codon. Heterozygosity for a naturally occurring single-nucleotide polymorphism (SNP) is evident downstream of the editing site.

**Extended Data Fig. 2 Editing of splice-site adenine bases throughout the human PCSK9 gene with first set of ten candidate gRNAs.**

Primary human hepatocytes were transfected at a dose of 5,000 ng RNA per ml; the boldface underlined letter in each of the following protospacer/PAM sequences (in which the solidus indicates the division between the protospacer and PAM) indicates the target splice-site adenine base. The black box in each panel indicates the gRNA protospacer sequence. All panels were generated with CRISPResso2. **a**, *PCSK9-1*, CCCGCA**A**CCTTGGCGCAGCGG/TGG. **b**, *PCSK9-2*, GGTGGCT**C**ACCAGCTCCAGC/AGG. **c**, *PCSK9-3*, GCTT**A**CCTGTCTGTGGAAGC/GGG. **d**, *PCSK9-4*, TGCTT**A** CCTGTCTGTGGAAG/CGG. **e**, *PCSK9-5*, TTGGAA**A**GACGGAGGCAGGCC/TGG. **f**, *PCSK9-6*, GAA**A**GACGGAGGCAGCCTGG/TGG. **g**, *PCSK9-7*, TCCC**A**GGCCTGGAGTTATT/CGG. **h**, *PCSK9-8*, AGCAC**C**CTCGGGAGCTG/AGG. **i**, *PCSK9-9*, CTTTCC**A**GGTCATCACAGTT/GGG. **j**, *PCSK9-10*, CCTTTCC**A**GGTCATCACAGT/TGG.

**Extended Data Fig. 3 Editing of splice-site adenine bases throughout the human PCSK9 gene with second set of ten candidate gRNAs.**

Primary human hepatocytes were transfected at a dose of 5,000 ng RNA per ml; the boldface underlined letter in each of the following protospacer/PAM sequences (in which the solidus indicates the division between the protospacer and PAM) indicates the target splice-site adenine base. The black box in each panel indicates the gRNA protospacer sequence. All panels were generated with CRISPResso2. **a**, *PCSK9-11*, TTTCC**A**GGTCATCACAGTTG/GGG. **b**, *PCSK9-12*, CTT**A** CCTGCCCATGGGTGC/TGG. **c**, *PCSK9-13*,

TAAG GCCCAAGGGGGCAAGC/TGG. **d**, *PCSK9-14*, CCTCTTCACCCTGCTCCTGAG/GGG. **e**, *PCSK9-15*, GCCTCTTCACCCTGCTCCTGA/GGG. **f**, *PCSK9-16*, TTCACCCTGCTCCTGAGGGGC/CGG. **g**, *PCSK9-17*, TACCCTGCTCCTGAGGGGCC/GGG. **h**, *PCSK9-18*, CCCAGGCTGCAGCTCCC ACT/GGG. **i**, *PCSK9-19*, CCCCAGGCTGCAGCTCCCAC/TGG. **j**, *PCSK9-20*, GCAGGTGACCGTGGCCTGCG/AGG.

#### Extended Data Fig. 4 Editing of PCSK9 exon 1 splice-donor adenine base in mice and in cynomolgus monkeys.

**a–f**, Representative liver samples from mice treated with LNPs with *PCSK9-1m* gRNA (portrayed in Fig. 1e). Protospacer/PAM sequence, CCCATACCTTGGAGCAACGG/CGG (in which the solidus indicates the division between the protospacer and PAM, and the boldface underlined letter indicates the target splice-donor adenine base). The black box in each panel indicates the gRNA protospacer sequence. All panels were generated with CRISPResso2. LNP doses were 2.0 mg kg<sup>-1</sup> (**a**), 1.0 mg kg<sup>-1</sup> (**b**), 0.5 mg kg<sup>-1</sup> (**c**), 0.25 mg kg<sup>-1</sup> (**d**), 0.125 mg kg<sup>-1</sup> (**e**) and 0.05 mg kg<sup>-1</sup> (**f**). **g–i**, Representative liver samples from three monkeys treated with a dose of 1.0 mg kg<sup>-1</sup> of LNPs with *PCSK9-1* gRNA, portrayed in Fig. 2a–d (treated monkeys 1, 2 and 3). Protospacer/PAM sequence, CCCGCAACTTGGCGCAGCGG/TGG (in which the solidus indicates the division between the protospacer and PAM, and the boldface underlined letter indicates the target splice-donor adenine base). The black box in each panel indicates the gRNA protospacer sequence. All panels were generated with CRISPResso2.

#### Extended Data Fig. 5 Liver function tests in short-term cynomolgus monkey studies.

**a**, Absolute values of blood levels of AST, ALT, alkaline phosphatase, γ-glutamyltransferase, total bilirubin and albumin in the three LNP-treated monkeys in Fig. 2a–d, as well as a contemporaneous control monkey that received PBS, at various time points up to 15 days. *n* = 1 blood sample per

monkey at each time point. Some values are missing for the control monkey (all day 3 values, all later  $\gamma$ -glutamyltransferase values). **b–g**, Absolute values of blood levels of AST (**b**), ALT (**c**), alkaline phosphatase (**d**),  $\gamma$ -glutamyltransferase (**e**), total bilirubin (**f**) and albumin (**g**) in the individual monkeys portrayed in Fig. [2e–g](#), as well as in non-contemporaneous control monkeys that received PBS, at various time points up to 15 days.  $n = 1$  blood sample per monkey at each time point. [Source data](#)

### [Extended Data Fig. 6 Pharmacokinetics of ABE8.8 and PCSK9-1 LNPs in cynomolgus monkeys.](#)

**a**, Plasma levels of ionizable cationic lipid and PEG-lipid components of ABE8.8 and *PCSK9-1* LNPs at various LNP doses in the monkeys portrayed in Fig. [2e–g](#) (mean  $\pm$  s.d. for each group,  $n = 3$  monkeys per dose group) at various time points up to 2 weeks after treatment. llod, lower limit of detection. **b**, Liver ABE8.8 mRNA levels in monkeys that received a dose of  $1.0 \text{ mg kg}^{-1}$  LNPs (mean  $\pm$  s.d. for each group,  $n = 4$  monkeys per necropsy group) at various time points up to 2 weeks after treatment.

[Source data](#)

### [Extended Data Fig. 7 Long-term effects of adenine base editing of PCSK9 on lipoprotein\(a\) in cynomolgus monkeys.](#)

Changes in the blood lipoprotein(a) level in the six monkeys from Fig. [3a](#), comparing levels at various time points up to 238 days after treatment versus the baseline level before treatment. Mean  $\pm$  s.d. for the LNP-treated group ( $n = 4$  monkeys) and mean for the control group ( $n = 2$  monkeys) at each time point. The dotted lines indicate 100% and 65% of baseline levels. [Source data](#)

### [Extended Data Fig. 8 Long-term pharmacodynamic effects of adenine base editing of PCSK9 in cynomolgus monkeys.](#)

**a–j**, Absolute values of blood levels of LDL cholesterol (**a**), total cholesterol (**b**), high-density lipoprotein (HDL) cholesterol (**c**), triglycerides (**d**), AST (**e**), ALT (**f**), alkaline phosphatase (**g**),  $\gamma$ -glutamyltransferase (**h**),

total bilirubin (**i**) and albumin (**j**) in the individual monkeys portrayed in Fig. 3 ( $n = 4$  monkeys treated with a dose of  $3.0 \text{ mg kg}^{-1}$  of an LNP formulation with ABE8.8 mRNA and *PCSK9-1* gRNA, and  $n = 2$  monkeys treated with PBS) at various time points up to 238 days after treatment. Shades of red represent LNP-treated monkeys, and shades of grey represent control monkeys. [Source data](#)

## [Extended Data Fig. 9 Additional studies with cynomolgus monkeys receiving a dose of \$3.0 \text{ mg kg}^{-1}\$ of LNPs.](#)

Levels of liver editing of the *PCSK9* exon 1 splice-donor adenine base (at day 15), blood AST and blood ALT.  $n = 3$  monkeys treated with PBS,  $n = 4$  monkeys treated with a dose of  $3.0 \text{ mg kg}^{-1}$  LNPs containing ABE8.8 mRNA and non-*PCSK9*-targeting gRNA and  $n = 4$  monkeys treated with a dose of  $3.0 \text{ mg kg}^{-1}$  ABE8.8 and *PCSK9-1* LNPs. Bar indicates mean value at each time point. [Source data](#)

## Supplementary information

### [Supplementary Tables](#)

This file contains Supplementary Tables 1-6.

### [Reporting Summary](#)

### [Supplementary Data](#)

This zipped file contains 3 files: **digenome\_step1.sh** - custom code for step 1 of the Digenome-seq analysis described in this manuscript; **digenome\_step2.R** custom code for step 2 of the Digenome-seq analysis described in this manuscript and a **README.txt** with instructions on how to run the custom code files.

## Source data

[\*\*Source Data Fig. 1\*\*](#)

[\*\*Source Data Fig. 2\*\*](#)

[\*\*Source Data Fig. 3\*\*](#)

[\*\*Source Data Fig. 4\*\*](#)

[\*\*Source Data Extended Data Fig. 5\*\*](#)

[\*\*Source Data Extended Data Fig. 6\*\*](#)

[\*\*Source Data Extended Data Fig. 7\*\*](#)

[\*\*Source Data Extended Data Fig. 8\*\*](#)

[\*\*Source Data Extended Data Fig. 9\*\*](#)

## Rights and permissions

[Reprints and Permissions](#)

## About this article



Check for  
updates

## Cite this article

Musunuru, K., Chadwick, A.C., Mizoguchi, T. *et al.* In vivo CRISPR base editing of *PCSK9* durably lowers cholesterol in primates. *Nature* **593**, 429–434 (2021). <https://doi.org/10.1038/s41586-021-03534-y>

## Download citation

- Received: 06 December 2020
- Accepted: 11 April 2021
- Published: 19 May 2021
- Issue Date: 20 May 2021
- DOI: <https://doi.org/10.1038/s41586-021-03534-y>

## Comments

By submitting a comment you agree to abide by our [Terms](#) and [Community Guidelines](#). If you find something abusive or that does not comply with our terms or guidelines please flag it as inappropriate.

[Access through your institution](#)

[Change institution](#)

[Buy or subscribe](#)

Advertisement

---

This article was downloaded by **calibre** from <https://www.nature.com/articles/s41586-021-03534-y>

- Article
- [Published: 05 May 2021](#)

# Distinct fission signatures predict mitochondrial degradation or biogenesis

- [Tatjana Kleele](#) [ORCID: orcid.org/0000-0001-5320-7337<sup>1</sup>](#),
- [Timo Rey](#) [ORCID: orcid.org/0000-0002-2124-937X<sup>1</sup>](#),
- [Julius Winter](#) [ORCID: orcid.org/0000-0001-6702-1330<sup>1</sup>](#),
- [Sofia Zaganelli](#) [ORCID: orcid.org/0000-0002-2815-3231<sup>1</sup>](#),
- [Dora Mahecic](#) [ORCID: orcid.org/0000-0003-1004-4771<sup>1</sup>](#),
- [Hélène Perreten Lambert<sup>1</sup>](#),
- [Francesco Paolo Ruberto<sup>2</sup>](#),
- [Mohamed Nemir<sup>2</sup>](#),
- [Timothy Wai<sup>3</sup>](#),
- [Thierry Pedrazzini](#) [ORCID: orcid.org/0000-0002-7969-8725<sup>2</sup>](#) &
- [Suliana Manley](#) [ORCID: orcid.org/0000-0002-4755-4778<sup>1</sup>](#)

[Nature](#) volume 593, pages 435–439 (2021) [Cite this article](#)

- 12k Accesses
- 394 Altmetric
- [Metrics details](#)

## Subjects

- [Fluorescence imaging](#)
- [Membrane biophysics](#)

- [Mitochondria](#)
- [Mitophagy](#)

## Abstract

Mitochondrial fission is a highly regulated process that, when disrupted, can alter metabolism, proliferation and apoptosis<sup>1,2,3</sup>. Dysregulation has been linked to neurodegeneration<sup>3,4</sup>, cardiovascular disease<sup>3</sup> and cancer<sup>5</sup>. Key components of the fission machinery include the endoplasmic reticulum<sup>6</sup> and actin<sup>7</sup>, which initiate constriction before dynamin-related protein 1 (DRP1)<sup>8</sup> binds to the outer mitochondrial membrane via adaptor proteins<sup>9,10,11</sup>, to drive scission<sup>12</sup>. In the mitochondrial life cycle, fission enables both biogenesis of new mitochondria and clearance of dysfunctional mitochondria through mitophagy<sup>1,13</sup>. Current models of fission regulation cannot explain how those dual fates are decided. However, uncovering fate determinants is challenging, as fission is unpredictable, and mitochondrial morphology is heterogeneous, with ultrastructural features that are below the diffraction limit. Here, we used live-cell structured illumination microscopy to capture mitochondrial dynamics. By analysing hundreds of fissions in African green monkey Cos-7 cells and mouse cardiomyocytes, we discovered two functionally and mechanistically distinct types of fission. Division at the periphery enables damaged material to be shed into smaller mitochondria destined for mitophagy, whereas division at the midzone leads to the proliferation of mitochondria. Both types are mediated by DRP1, but endoplasmic reticulum- and actin-mediated pre-constriction and the adaptor MFF govern only midzone fission. Peripheral fission is preceded by lysosomal contact and is regulated by the mitochondrial outer membrane protein FIS1. These distinct molecular mechanisms explain how cells independently regulate fission, leading to distinct mitochondrial fates.

## Access options

Subscribe to Journal

Get full journal access for 1 year

\$199.00

only \$3.90 per issue

[Subscribe](#)

All prices are NET prices.

VAT will be added later in the checkout.

Tax calculation will be finalised during checkout.

Rent or Buy article

Get time limited or full article access on ReadCube.

from \$8.99

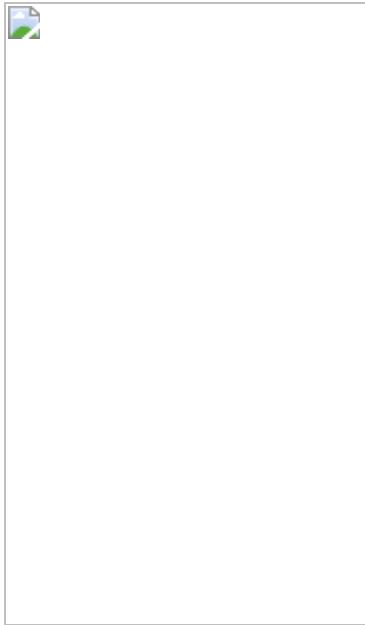
[Rent or Buy](#)

All prices are NET prices.

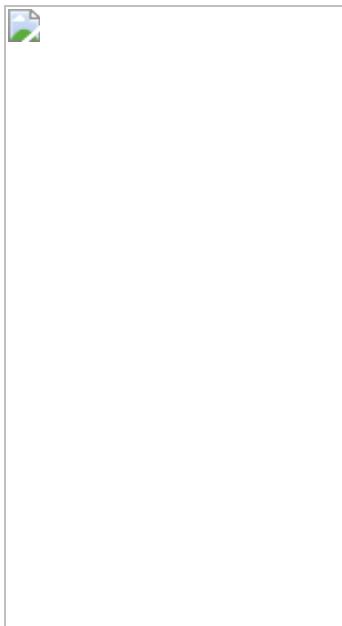
### **Additional access options:**

- [Log in](#)
- [Access through your institution](#)
- [Learn about institutional subscriptions](#)

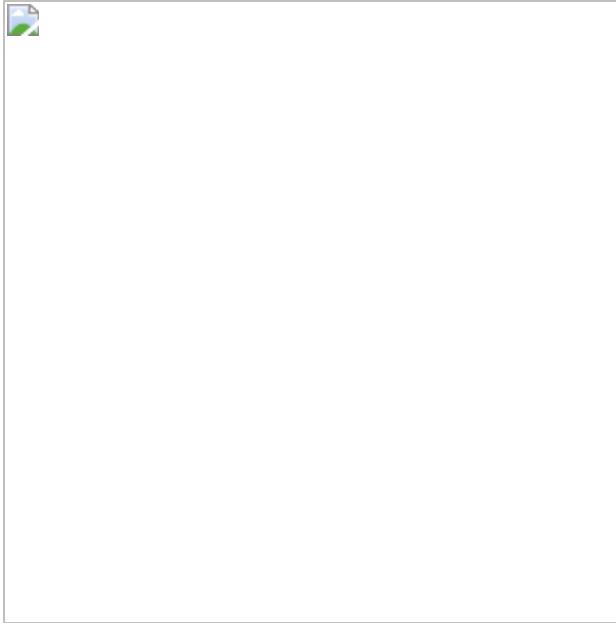
**Fig. 1: Mitochondrial fissions are bimodally positioned and linked to distinct physiologies.**



**Fig. 2: Midzone and peripheral fissions differ in mitochondrial DNA content and fates.**



**Fig. 3: Midzone and peripheral fissions are independently regulated by distinct molecular machineries.**



## Data availability

All imaging as well as numerical data relevant to this study are publicly available on the online repository Zenodo (<https://doi.org/10.5281/zenodo.3550643>). The data are organized according to their appearance in the main figures and Extended Data figures. The unprocessed western blot gels are provided in Supplementary Fig. 1. Plasmids and cell lines are available from the corresponding authors on request. [Source data](#) are provided with this paper.

## Code availability

The custom-written Fiji-macro script for BrU quantification is available at <https://github.com/TimoHenry/MitochondrialRNAgranules> and the custom written Fiji-macro for caspase staining is available at [https://github.com/jutziw/mitochondrial\\_division](https://github.com/jutziw/mitochondrial_division).

## References

1. 1.

Twig, G. et al. Fission and selective fusion govern mitochondrial segregation and elimination by autophagy. *EMBO J.* **27**, 433–446 (2008).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

2. 2.

Rambold, A. S., Kostelecky, B., Elia, N. & Lippincott-Schwartz, J. Tubular network formation protects mitochondria from autophagosomal degradation during nutrient starvation. *Proc. Natl Acad. Sci. USA* **108**, 10190–10195 (2011).

[ADS](#) [CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

3. 3.

Song, M., Franco, A., Fleischer, J. A., Zhang, L. & Dorn, G. W., II. Abrogating mitochondrial dynamics in mouse hearts accelerates mitochondrial senescence. *Cell Metab.* **26**, 872–883 (2017).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

4. 4.

Wang, W. et al. Parkinson’s disease-associated mutant VPS35 causes mitochondrial dysfunction by recycling DLP1 complexes. *Nat. Med.* **22**, 54–63 (2016).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

5. 5.

Ma, J.-T. et al. Effects of dynamin-related protein 1 regulated mitochondrial dynamic changes on invasion and metastasis of lung cancer cells. *J. Cancer* **10**, 4045–4053 (2019).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

6. 6.

Friedman, J. R. et al. ER tubules mark sites of mitochondrial division. *Science* **334**, 358–362 (2011).

[ADS](#) [CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

7. 7.

Ji, W. K., Hatch, A. L., Merrill, R. A., Strack, S. & Higgs, H. N. Actin filaments target the oligomeric maturation of the dynamin GTPase Drp1 to mitochondrial fission sites. *eLife* **4**, e11553 (2015).

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

8. 8.

Smirnova, E., Griparic, L., Shurland, D. L. & van der Bliek, A. M. Dynamin-related protein Drp1 is required for mitochondrial division in mammalian cells. *Mol. Biol. Cell* **12**, 2245–2256 (2001).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

9. 9.

Gandre-Babbe, S. & van der Bliek, A. M. The novel tail-anchored membrane protein Mff controls mitochondrial and peroxisomal fission in mammalian cells. *Mol. Biol. Cell* **19**, 2402–2412 (2008).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

10. 10.

Palmer, C. S. et al. MiD49 and MiD51, new components of the mitochondrial fission machinery. *EMBO Rep.* **12**, 565–573 (2011).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

11. 11.

Osellame, L. D. et al. Cooperative and independent roles of the Drp1 adaptors Mff, MiD49 and MiD51 in mitochondrial fission. *J. Cell Sci.* **129**, 2170–2181 (2016).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

12. 12.

Fröhlich, C. et al. Structural insights into oligomerization and mitochondrial remodelling of dynamin 1-like protein. *EMBO J.* **32**, 1280–1292 (2013).

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

13. 13.

Burman, J. L. et al. Mitochondrial fission facilitates the selective mitophagy of protein aggregates. *J. Cell Biol.* **216**, 3231–3247 (2017).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

14. 14.

Ashrafi, G. & Schwarz, T. L. The pathways of mitophagy for quality control and clearance of mitochondria. *Cell Death Differ.* **20**, 31–42 (2013).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

15. 15.

Sturtz, L. A., Diekert, K., Jensen, L. T., Lill, R. & Culotta, V. C. A fraction of yeast Cu,Zn-superoxide dismutase and its metallochaperone, CCS, localize to the intermembrane space of mitochondria. A physiological role for SOD1 in guarding against mitochondrial oxidative damage. *J. Biol. Chem.* **276**, 38084–38089 (2001).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

16. 16.

Mallilankaraman, K. et al. MICU1 is an essential gatekeeper for MCU-mediated mitochondrial Ca<sup>2+</sup> uptake that regulates cell survival. *Cell* **151**, 630–644 (2012).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

17. 17.

Jourdain, A. A. et al. GRSF1 regulates RNA processing in mitochondrial RNA granules. *Cell Metab.* **17**, 399–410 (2013).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

18. 18.

Wong, Y. C., Ysselstein, D. & Krainc, D. Mitochondria-lysosome contacts regulate mitochondrial fission via RAB7 GTP hydrolysis. *Nature* **554**, 382–386 (2018).

[ADS](#) [CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

19. 19.

Soubannier, V., Rippstein, P., Kaufman, B. A., Shoubridge, E. A. & McBride, H. M. Reconstitution of mitochondria derived vesicle formation demonstrates selective enrichment of oxidized cargo. *PLoS One* **7**, e52830 (2012).

[ADS](#) [CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

20. 20.

Pickles, S., Vigié, P. & Youle, R. J. Mitophagy and quality control mechanisms in mitochondrial maintenance. *Curr. Biol.* **28**, R170–R185 (2018).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

21. 21.

Korobova, F., Ramabhadran, V. & Higgs, H. N. An actin-dependent step in mitochondrial fission mediated by the ER-associated formin INF2. *Science* **339**, 464–467 (2013).

[ADS](#) [CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

22. 22.

Hirabayashi, Y. et al. ER-mitochondria tethering by PDZD8 regulates Ca<sup>2+</sup> dynamics in mammalian neurons. *Science* **358**, 623–630 (2017).

[ADS](#) [CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

23. 23.

Murley, A. et al. ER-associated mitochondrial division links the distribution of mitochondria and mitochondrial DNA in yeast. *eLife* **2**, e00422 (2013).

[ADS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

24. 24.

Lewis, S. C., Uchiyama, L. F. & Nunnari, J. ER-mitochondria contacts couple mtDNA synthesis with mitochondrial division in human cells. *Science* **353**, aaf5549 (2016).

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

25. 25.

Yu, R., Jin, S.-B., Lendahl, U., Nistér, M. & Zhao, J. Human Fis1 regulates mitochondrial dynamics through inhibition of the fusion machinery. *EMBO J.* **38**, e99748 (2019).

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

26. 26.

Xian, H., Yang, Q., Xiao, L., Shen, H.-M. & Liou, Y.-C. STX17 dynamically regulated by Fis1 induces mitophagy via hierarchical macroautophagic mechanism. *Nat. Commun.* **10**, 2059 (2019).

[ADS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

27. 27.

Kimura, W. et al. Redox signaling in cardiac renewal. *Antioxid. Redox Signal.* **21**, 1660–1673 (2014).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

28. 28.

Chang, F. & Minc, N. Electrochemical control of cell and tissue polarity. *Annu. Rev. Cell Dev. Biol.* **30**, 317–336 (2014).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

29. 29.

Wolf, D. M. et al. Individual cristae within the same mitochondrion display different membrane potentials and are functionally independent. *EMBO J.* **38**, e101056 (2019).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

30. 30.

Koirala, S. et al. Interchangeable adaptors regulate mitochondrial dynamin assembly for membrane scission. *Proc. Natl Acad. Sci. USA* **110**, E1342–E1351 (2013).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

31. 31.

Mozdy, A. D., McCaffery, J. M. & Shaw, J. M. Dnm1p GTPase-mediated mitochondrial fission is a multi-step process requiring the novel integral membrane component Fis1p. *J. Cell Biol.* **151**, 367–380 (2000).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

32. 32.

Song, W. et al. Mutant huntingtin binds the mitochondrial fission GTPase dynamin-related protein-1 and increases its enzymatic activity. *Nat. Med.* **17**, 377–382 (2011).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

33. 33.

Guo, X. et al. Inhibition of mitochondrial fragmentation diminishes Huntington’s disease-associated neurodegeneration. *J. Clin. Invest.* **123**, 5371–5388 (2013).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

34. 34.

Zamora, M., Pardo, R. & Villena, J. A. Pharmacological induction of mitochondrial biogenesis as a therapeutic strategy for the treatment of type 2 diabetes. *Biochem. Pharmacol.* **98**, 16–28 (2015).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

35. 35.

Poburko, D., Santo-Domingo, J. & Demaurex, N. Dynamic regulation of the mitochondrial proton gradient during cytosolic calcium elevations. *J. Biol. Chem.* **286**, 11672–11684 (2011).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

36. 36.

Gutscher, M. et al. Real-time imaging of the intracellular glutathione redox potential. *Nat. Methods* **5**, 553–559 (2008).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

37. 37.

Suzuki, J. et al. Imaging intraorganellar  $\text{Ca}^{2+}$  at subcellular resolution using CEPIA. *Nat. Commun.* **5**, 4153 (2014).

[ADS](#) [CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

38. 38.

Minin, A. A. et al. Regulation of mitochondria distribution by RhoA and formins. *J. Cell Sci.* **119**, 659–670 (2006).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

39. 39.

Narendra, D., Tanaka, A., Suen, D.-F. & Youle, R. J. Parkin is recruited selectively to impaired mitochondria and promotes their autophagy. *J. Cell Biol.* **183**, 795–803 (2008).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

40. 40.

Jackson, W. T. et al. Subversion of cellular autophagosomal machinery by RNA viruses. *PLoS Biol.* **3**, e156 (2005).

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

41. 41.

Wu, J. et al. Improved orange and red Ca<sup>2+</sup> indicators and photophysical considerations for optogenetic applications. *ACS Chem. Neurosci.* **4**, 963–972 (2013).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

42. 42.

Spelbrink, J. N. et al. Human mitochondrial DNA deletions associated with mutations in the gene encoding Twinkle, a phage T7 gene 4-like protein localized in mitochondria. *Nat. Genet.* **28**, 223–231 (2001).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

43. 43.

Rey, T. et al. Mitochondrial RNA granules are fluid condensates positioned by membrane dynamics. *Nat. Cell Biol.* **22**, 1180–1186 (2020).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

44. 44.

Ran, F. A. et al. Genome engineering using the CRISPR–Cas9 system. *Nat. Protocols* **8**, 2281–2308 (2013).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

45. 45.

Marroquin, L. D., Hynes, J., Dykens, J. A., Jamieson, J. D. & Will, Y. Circumventing the Crabtree effect: replacing media glucose with galactose increases susceptibility of HepG2 cells to mitochondrial toxicants. *Toxicol. Sci.* **97**, 539–547 (2007).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

46. 46.

Eulalio, A. et al. Functional screening identifies miRNAs inducing cardiac regeneration. *Nature* **492**, 376–381 (2012).

[ADS](#) [CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

47. 47.

Kelso, G. F. et al. Selective targeting of a redox-active ubiquinone to mitochondria within cells: antioxidant and antiapoptotic properties. *J. Biol. Chem.* **276**, 4588–4596 (2001).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

48. 48.

York, A. G. et al. Instant super-resolution imaging in live cells and embryos via analog image processing. *Nat. Methods* **10**, 1122–1126 (2013).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

49. 49.

Mahecic, D. et al. Homogeneous multifocal excitation for high-throughput super-resolution imaging. *Nat. Methods* **17**, 726–733 (2020).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

50. 50.

Holmes, T. J. & Liu, Y. H. Richardson-Lucy/maximum likelihood image restoration algorithm for fluorescence microscopy: further testing. *Appl. Opt.* **28**, 4930–4938 (1989).

[ADS](#) [CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

51. 51.

Schindelin, J. et al. Fiji: an open-source platform for biological-image analysis. *Nat. Methods* **9**, 676–682 (2012).

[CAS](#) [Google Scholar](#)

52. 52.

Schneider, C. A., Rasband, W. S. & Eliceiri, K. W. NIH Image to ImageJ: 25 years of image analysis. *Nat. Methods* **9**, 671–675 (2012).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

[Download references](#)

## Acknowledgements

We thank C. Cottiny and M. Colomer for experimental and technical assistance, M.-C. Croisier and G. Knott from the BioEM (EPFL) for carrying out the electron microscopy, T. Laroche (BIOP, EPFL) for imaging support and M. Ryan for the MEF lines. We thank T. Misgeld, J.-C. Martinou and P. Ramdy for feedback on the manuscript. Research in S.M.’s laboratory is supported by the National Centre of Competence in Research Chemical Biology, the European Research Council (CoG 819823, Piko) and the Swiss National Science Foundation (182429). T.K. received funding from the European Molecular Biology Organization (ALTF-739-2016) and the Munich Cluster for Systems Neurology (SyNergy). This work is supported in part by grants from the Swiss National Science Foundation to T.P. (no. CRSII5\_173738 and no. 31003A\_182322).

## Author information

### Affiliations

1. Institute of Physics, École Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland

Tatjana Kleele, Timo Rey, Julius Winter, Sofia Zaganelli, Dora Mahecic, Hélène Perreten Lambert & Suliana Manley

2. Experimental Cardiology Unit, Department of Cardiovascular Medicine, University of Lausanne Medical School, Lausanne, Switzerland

Francesco Paolo Ruberto, Mohamed Nemir & Thierry Pedrazzini

3. Mitochondrial Biology Group, Institut Pasteur, CNRS UMR 3691, Paris, France

Timothy Wai

## Authors

1. Tatjana Kleele

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

2. Timo Rey

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

3. Julius Winter

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

4. Sofia Zaganelli

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

5. Dora Mahecic

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

6. Hélène Perreten Lambert  
[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

7. Francesco Paolo Ruberto  
[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

8. Mohamed Nemir  
[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

9. Timothy Wai  
[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

10. Thierry Pedrazzini  
[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

11. Suliana Manley  
[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

## Contributions

T.K. and S.M. conceived the project and designed experiments. T.K. performed imaging experiments and analysis. T.R. performed the TWINKLE, FASTKD2, BrU and MitoSOX imaging and contributed to the analysis. T.K. and J.W. performed the caspase, DRP1 and LC3 imaging. D.M. developed and adjusted the iSIM set-up. T.K., T.R. and D.M.

performed the transmission electron microscopy (TEM) experiments. S.Z. performed the western blots. F.P.R., M.N. and T.P. designed and performed culturing of mouse cardiomyocytes and the proliferation assay. T.W., S.Z. and H.P.L. designed and cloned the CRISPR–Cas9 transgenic lines. T.K. and S.M. designed figures and wrote the manuscript, with input from all authors.

## Corresponding authors

Correspondence to [Tatjana Kleele](#) or [Suliana Manley](#).

## Ethics declarations

## Competing interests

The authors declare no competing interests.

## Additional information

**Peer review information** *Nature* thanks Henry Higgs and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

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

## Extended data figures and tables

### [Extended Data Fig. 1 Distribution of mitochondrial fission sites.](#)

**a**, Histogram of the relative position of constriction/fission measured by mitochondrial volume ( $n = 190$  fissions). The two peaks are coloured in orange (0–25 bin; ‘peripheral position’) and green (25–50 bin; ‘midzone position’). **b**, Relative position of mitochondrial fission measured by length

versus measured by area ( $n = 190$  fissions). **c**, Stacked histogram of the relative position of fission for different bins grouped by the total length of the dividing mitochondria ( $n = 1393$ ). **d**, Scatter plot of the total length of dividing mitochondria versus the relative position of the fission site along the length axis with peripheral fissions (0–25% bin) coloured in orange and midzone fissions (25–50% bin) coloured in green ( $n = 1393$  fissions). **e**, Length distribution of the smaller (light colour) and larger (dark colour) daughter mitochondria arising from peripheral left, orange) and midzone (right, green) fissions. **f**, Histogram of relative position of fission in datasets acquired with a mitochondrial inner membrane marker (left, COX8 targeting domain;  $n = 510$  fissions) and a mitochondrial outer membrane marker (right, TOM20;  $n = 368$  fissions). [Source data](#)

## Extended Data Fig. 2 Physiological changes preceding fission are independent of volume and absolute length.

**a**, Normalized mito–GFP intensity depending on the relative position of fission measured in mitochondria immediately before fission. ( $n = 50$  fissions). **b**, Dependence on the length of the daughter mitochondria of normalized mito–GFP intensity immediately before peripheral (orange) or midzone (green) fissions ( $n = 50$  fissions). **c**, Dependence on the length of the daughter mitochondria of normalized TMRE intensity immediately before peripheral (orange) or midzone (green) fissions ( $n = 56$  fissions). **d**, Dependence on the length of the daughter mitochondria of normalized mito–SypHer intensity before fission ( $n = 53$  fissions). **e**, Dependence of the length of the daughter mitochondria on normalized MitoSOX intensity before fission ( $n = 52$  fissions). **f**, Rates of peripheral and midzone fissions in control Cos-7 cells ( $n = 10$  fields of view) versus cells treated with 500 nM ROS scavenger MitoQ ( $n = 10$  fields of view). **g**, Dependence on the length of the daughter mitochondria of ratiometric intensity of mito–GRX1–roGFP immediately before fission ( $n = 52$  fissions). **h**, Dependence on the length of the daughter mitochondria of normalized mito–R–GECO1 intensity before fission ( $n = 50$  fissions). **i**, Dependence on the length of the daughter mitochondria of normalized CEPIA3–mt intensity before fission ( $n = 61$  fissions). **j**, Mitochondrial pH before and after a peripheral or midzone fission from SIM movies of mito–SypHer transfected Cos-7 cells. **k**, Normalized mito–SypHer intensity as a function of relative position of

fission, measured immediately before fission ( $n = 53$  fissions). **l**, Mitochondrial ROS before and after a peripheral or midzone fission from ratiometric images of mito–GRX1–roGFP transfected Cos-7 cells. **m**, Ratiometric mito–GRX1–roGFP intensity as a function of relative position of fission, measured immediately before fission ( $n = 52$  fissions). **n**, Mitochondrial matrix  $\text{Ca}^{2+}$  before and after a peripheral or midzone fission from SIM movies of mito–R–GECO1 transfected Cos-7 cells. **o**, Normalized mito–R–GECO1 intensity as a function of relative position of fission, measured immediately before fission ( $n = 50$  fissions). **p**, Mitochondrial matrix  $\text{Ca}^{2+}$  before and after a peripheral or midzone fission from SIM movies of CEPIA3–mt transfected Cos-7 cells. **q**, Normalized CEPIA3–mt intensity as a function of relative position of fission, measured immediately before fission ( $n = 61$  fissions). In **a**, **k**, **m**, **o** and **q** circles indicate individual measurements; values of binned groups represented as box plots (line, mean; bounds of box: 25th and 75th percentiles; whiskers, min/max values). Light blue areas indicate mean intensity in non-dividing mitochondria ( $\pm$  s.d.). n.s.  $>0.05$ , \*\*\* $P < 0.001$ . Number of experiments, statistical tests and exact  $P$  values are provided in Supplementary Table 1. Scale bars, 0.5  $\mu\text{m}$ . Fission sites are indicated by arrowheads. [Source data](#)

### Extended Data Fig. 3 Redistribution of mitochondrial DNA and RNA granules.

**a**, Distribution of PicoGreen foci in the small and large daughter mitochondrion derived from peripheral and midzone fission ( $n = 78$  fissions). **b**, Normalized TMRE intensity in the small daughter mitochondria from peripheral fissions that contain 0, 1 or 2 nucleoids ( $n = 20$  fissions). **c**, SIM images of mitochondrial RNA granules (MRGs) (FASTKD2) before and after fission. **d**, Number of MRGs per  $\mu\text{m}$  length as a function of fission position ( $n = 84$  fissions). Blue line shows average MRG per length in non-dividing mitochondria ( $n = 41$ ). **e**, Distribution of the number of MRGs (FASTKD2;  $n = 84$  fissions). **f**, Number of replicating nucleoids (TWINKLE;  $n = 74$  fissions) in smaller (light) and larger (dark) daughter mitochondria from peripheral (orange) and midzone (green). **g**, Number of BrU-positive foci per cell in control Cos-7 cells ( $n = 39$ ) and cells exposed to UV light for 3 min before measurement ( $n = 98$ ). **h**, Time-lapse SIM sequence of Cos-7 mitochondria (mito–RFP, grey) and

autophagosomes (EGFP–LC3, green), where the small daughter mitochondrion from a peripheral fission is being taken up by an autophagosome (asterisk). In **b** and **g** circles indicate individual measurements; line, mean; bounds of box, 25th and 75th percentiles; whiskers, min/max values. \* $P < 0.05$ , \*\*\* $P < 0.001$ . Number of experiments, statistical tests and exact  $P$  values are provided in Supplementary Table 1. Scale bars, 0.5  $\mu\text{m}$ . [Source data](#)

#### [Extended Data Fig. 4 Time course of physiological changes and recruitment of fission regulators.](#)

**a, b**, Time course of fluorescent signals in four examples of Cos-7 mitochondria displaying normalized TMRE intensity (**a**) and mito–R-GECO1 intensity (**b**) with corresponding SIM images in the mitochondrial compartment giving rise to the smaller daughter mitochondria before a peripheral division. **c**, Average inner membrane diameter at the constriction site at several time points before fission, measured in mito–R-GECO1 transfected Cos-7 cells during the time window where  $\text{Ca}^{2+}$  is elevated (green box in **b**,  $n = 10$  fission events). **d, e**, Time course of lysosome co-localization (**d**) and autophagosome co-localization (**e**) at constriction sites for peripheral fissions, by measuring LAMP1–mEGFP and EGFP–LC3 intensity, respectively. For EGFP–LC3 measurements, cells were pre-treated with 10  $\mu\text{M}$  CCCP to increase LC3 signals. **f**, Normalized DRP1 intensity at the constriction sites before peripheral fission in four examples of Cos-7 mitochondria with corresponding SIM images. Blue dotted lines ( $t = 0$  s) mark the time point of fission. [Source data](#)

#### [Extended Data Fig. 5 Peripheral and midzone fissions interact differently with the mitochondrial network, and the distribution of the fission positions is regulated independently.](#)

**a**, Schematic diagram depicting the fate ('no event', another 'fission' or 'fusion') of each daughter mitochondrion from peripheral and midzone fissions after the initial division in postnatal cardiomyocytes. Only mitochondria that could be traced for more than 100 s after fission were included in the analysis. **b**, Distribution of the relative position of fission in

starved Cos-7 cells, with peripheral (1–25%) fission labelled in orange and midzone fissions (25–50%) in green ( $n = 212$  fissions). The frequency distribution of Cos-7 control samples is superimposed in grey (replotted from Fig. 1c). **c, d**, Distribution of the relative position of constriction/fission along the length axis of isoproterenol-treated mouse cardiomyocyte mitochondria (**c**;  $n = 356$  fissions) and miR-199 treated cardiomyocytes (**d**;  $n = 225$  fissions) respectively. The frequency distribution of untreated mouse cardiomyocyte samples is superimposed in grey (replotted from Fig. 1e). [Source data](#)

## **Extended Data Fig. 6 Peripheral and midzone fissions are both DRP1 mediated but involve distinct upstream mechanisms.**

**a**, Two-colour SIM images of mitochondria (mito–GFP, greyscale) and DRP1 (mCherry–DRP1, red) undergoing peripheral or midzone fission. **b**, Normalized DRP1 intensity on the constriction sites of peripheral and midzone divisions. The threshold for a DRP1 accumulation (blue dotted line) was set at a signal  $>3\times$  over background ( $n = 107$  fissions). **c**, Time-lapse sequence of a SIM movie, in which both mitochondrial outer membrane (TOM20–GFP) and inner membrane (mito–RFP) were labelled to detect mitochondria-derived vesicle (MDV) formation (arrowhead). **d**, Quantification of the fission positions for mitochondria undergoing MDV formation or not before or after division ( $n = 41$  fissions). **e**, SIM images of peripheral and midzone constrictions in fixed Cos-7 cells labelled with anti-PDZD8 (red). **f**, Distribution of normalized fluorescence intensities of anti-PDZD8 staining in fixed Cos-7 cells for peripheral (orange) and midzone (green) fissions ( $n = 38$  fissions). **g**, Rate of peripheral and midzone fissions in control cells ( $n = 10$  FOV) versus cells treated with *INF2* siRNA ( $n = 10$  FOV). **h**, Correlated confocal and transmission electron microscopy (CLEM) of mitochondria in Cos-7 cells labelled with mito–GFP, fixed 24 h after expression. Zoom-in of two individual mitochondria with a peripheral (orange frame) and a midzone (green frame) constriction in the TEM plane containing the widest diameter of the constriction site. Pseudo-colouring of three consecutive TEM  $z$ -sections recombined into a single rendering shows mitochondria (green) and ER (magenta). Scale bar represents 2  $\mu\text{m}$  in confocal and 200 nm in TEM images. **i, k, m**, SIM images of peripheral and midzone constrictions in fixed Cos-7 cells labelled with anti-TOM20 (grey)

and anti-MFF (green) (**i**), anti-FIS1 (red) (**k**) and anti-MiD49 (green) (**m**). **j**, **l**, **n**, Distribution of normalized fluorescent intensities of anti-MFF ( $n = 92$  fissions) (**j**), anti-FIS1 ( $n = 59$  fissions) (**l**) and anti-MiD49 ( $n = 29$  fissions) (**n**) staining in fixed Cos-7 cells for peripheral (orange) and midzone (green) fissions. **o**, Normalized FIS1–GFP intensity at the fission site of peripheral (left) and midzone (right) divisions (orange dots: peripheral, green dots: midzone;  $n = 35$  fissions). **p**, Quantification of the peripheral (orange) and midzone (green) fission rates in wild-type ( $n = 16$  FOV) and *Mff*<sup>-/-</sup>*Mid49*<sup>-/-</sup>*Mid51*<sup>-/-</sup> triple-knockout ( $n = 16$  FOV) MEFs stained with Mitotracker Green. **q**, Quantification of the peripheral (orange) and midzone (green) fission rates in control Cos-7 cells and cells treated with siRNAs against *MFF* or *FIS1* ( $n \geq 13$  FOV per group). n.s.  $P > 0.05$ , \*\* $P < 0.01$ , \*\*\* $P < 0.001$ . Number of experiments, statistical tests and exact  $P$  values are provided in Supplementary Table 1. Scale bars, 0.5  $\mu\text{m}$ . Fission sites are indicated by arrowheads. [Source data](#)

### Extended Data Fig. 7 Silencing efficiency of siRNA against *FIS1*, *MFF* and *INF2* in Cos-7 cells.

**a**, Western blot analysis of protein levels for Cos-7 cells 72 h after transfection with two siRNAs against *INF2* at two different quantities (pmol). Molecular mass is in kilodaltons. **b**, Western blot analysis of protein levels for Cos-7 cells 72 h after transfection with 12.5 pmol of siRNA against *FIS1* or *MFF*. **c**, **d**, Mean fluorescence intensity in U2OS cells endogenously expressing MFF–GFP (**c**) or FIS1–GFP (**d**) in control conditions and 72 h after transfection with siRNA against *FIS1* or *MFF* ( $n > 39$  cells per group). **e**, Relative *Mff* RNA expression levels in control cardiomyocytes and in cardiomyocytes 48 h after transfection with si*Mff* and with or without isoproterenol (Iso) treatment. **f**, Relative *Fis1* RNA expression levels in control cardiomyocytes and in cardiomyocytes 48 h after transfection with si*Fis1* and with or without isoproterenol (Iso) treatment. \* $P < 0.05$ , \*\* $P < 0.01$ , \*\*\* $P < 0.001$ . Number of experiments, statistical tests and exact  $P$  values are provided in Supplementary Table 1. [Source data](#)

## Supplementary information

## **Supplementary Figure 1**

Uncropped western blot images shown in Extended Data Fig. 7 a b. Black dashed boxes correspond to the regions cropped for the Extended Data Figure.

## **Reporting Summary**

### **Supplementary Table 1**

| Statistics and Reproducibility. Detailed information about sample sizes, biological replicates, statistical test and exact P-values and confidence intervals for all numerical data.

### **Supplementary Video 1 Live imaging of mitochondrial fissions.**

Live-cell SIM imaging of peripheral and midzone mitochondrial divisions in Cos-7 cells labelled with Mitotracker green. Video was acquired at 1 frame/sec for 3 minutes and corresponds to Fig. 1 b.

### **Supplementary Video 2 Live imaging of mouse cardiomyocyte.**

Live-cell iSIM imaging of a post-natal mouse cardiomyocyte labelled with Mitotracker green showing peripheral (orange arrowhead) and midzone (green arrowhead) fissions. Video was acquired at 1 frame/5 sec for 6 minutes. Video corresponds to Fig. 1 f.

### **Supplementary Video 3 Live imaging of mitochondrial membrane potential (TMRE).**

Live-cell SIM imaging of TMRE stained Cos-7 cells showing a peripheral mitochondrial divisions. Video was acquired at 1 frame/3 sec. To highlight differences in fluorescent intensities, a heat-map look-up table was chosen.

### **Supplementary Video 4 Live imaging of mitochondrial matrix pH (Mito-SyPHer).**

Live-cell SIM imaging of Cos-7 transfected with mito–SypHer showing a peripheral mitochondrial divisions. Video was acquired at 1 frame/1.5 sec. To highlight differences in fluorescent intensities, a heat-map look-up table was chosen.

**Supplementary Video 5 Live imaging of mitochondrial Ca<sup>2+</sup> (mito–R-GECO1).**

Live-cell SIM imaging of Cos-7 transfected with mito–R-GECO1 showing a peripheral mitochondrial divisions. Video was acquired at 1 frame/3 sec. To highlight differences in fluorescent intensities, a heat-map look-up table was chosen.

**Supplementary Video 6 Live imaging of mitochondria-lysosome contact during fission.**

Live-cell iSIM imaging of lysosomes (LAMP1–mEGFP, green) and mitochondria (mito–TagRFP, grey) in Cos-7 cells. During peripheral fission (left) but not during midzone fission (right), lysosomes contact the mitochondrial constriction site. Video was acquired at 1 frame/3 sec for 3 minutes. Video corresponds to Fig. 3 d.

**Supplementary Video 7 Live imaging of mitochondria-autophagosome contact after fission.**

Live-cell iSIM imaging of Cos-7 cells where autophagosomes (LC3–GFP, green) and mitochondria (mito–TagRFP, grey) are labelled. After peripheral fission, the small daughter mitochondrion is subsequently engulfed by an autophagosome. Video was acquired at 1 frame/5 sec for 6.5 minutes. Video corresponds to Fig. 3 h.

**Supplementary Video 8 Live imaging of mitochondria-ER contacts during fission.**

Live-cell iSIM imaging of Cos-7 cells transfected with KDEL–RFP (ER) and mito–GFP (mitochondria) showing a midzone fission (left, green

arrowhead) and a peripheral fission (right, orange arrowhead). Video was acquired at 1 frame/14 sec for 5 minutes.

## Source data

[Source Data Fig. 1](#)

[Source Data Fig. 2](#)

[Source Data Fig. 3](#)

[Source Data Extended Data Fig. 1](#)

[Source Data Extended Data Fig. 2](#)

[Source Data Extended Data Fig. 3](#)

[Source Data Extended Data Fig. 4](#)

[Source Data Extended Data Fig. 5](#)

[Source Data Extended Data Fig. 6](#)

[Source Data Extended Data Fig. 7](#)

## Rights and permissions

[Reprints and Permissions](#)

## About this article



Check for  
updates

## Cite this article

Kleele, T., Rey, T., Winter, J. *et al.* Distinct fission signatures predict mitochondrial degradation or biogenesis. *Nature* **593**, 435–439 (2021). <https://doi.org/10.1038/s41586-021-03510-6>

[Download citation](#)

- Received: 15 November 2019
- Accepted: 31 March 2021
- Published: 05 May 2021
- Issue Date: 20 May 2021
- DOI: <https://doi.org/10.1038/s41586-021-03510-6>

## Comments

By submitting a comment you agree to abide by our [Terms](#) and [Community Guidelines](#). If you find something abusive or that does not comply with our terms or guidelines please flag it as inappropriate.

[Access through your institution](#)

[Change institution](#)

[Buy or subscribe](#)

## Associated Content

[Revolutionary view of two ways to split a mitochondrion](#)

- Rajarshi Chakrabarti
- Henry N. Higgs

Nature News & Views 05 May 2021

## Advertisement

---

This article was downloaded by **calibre** from <https://www.nature.com/articles/s41586-021-03510-6>

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

- Article
- [Published: 25 March 2021](#)

# Neuronal enhancers are hotspots for DNA single-strand break repair

- [Wei Wu](#) ORCID: [orcid.org/0000-0001-5164-2765](https://orcid.org/0000-0001-5164-2765)<sup>1</sup> na1,
- [Sarah E. Hill](#) ORCID: [orcid.org/0000-0002-8499-7455](https://orcid.org/0000-0002-8499-7455)<sup>2</sup> na1,
- [William J. Nathan](#) ORCID: [orcid.org/0000-0003-2752-7630](https://orcid.org/0000-0003-2752-7630)<sup>1,3</sup> na1,
- [Jacob Paiano](#)<sup>1</sup>,
- [Elsa Callen](#)<sup>1</sup>,
- [Dongpeng Wang](#) ORCID: [orcid.org/0000-0002-8272-1228](https://orcid.org/0000-0002-8272-1228)<sup>1</sup>,
- [Kenta Shinoda](#)<sup>1</sup>,
- [Niek van Wietmarschen](#)<sup>1</sup>,
- [Jennifer M. Colón-Mercado](#)<sup>2</sup>,
- [Dali Zong](#)<sup>1</sup>,
- [Raffaella De Pace](#) ORCID: [orcid.org/0000-0002-8179-6869](https://orcid.org/0000-0002-8179-6869)<sup>4</sup>,
- [Han-Yu Shih](#) ORCID: [orcid.org/0000-0002-7525-5433](https://orcid.org/0000-0002-7525-5433)<sup>5</sup>,
- [Steve Coon](#) ORCID: [orcid.org/0000-0002-5233-0277](https://orcid.org/0000-0002-5233-0277)<sup>4</sup>,
- [Maia Parsadanian](#)<sup>2</sup>,
- [Raphael Pavani](#)<sup>1</sup>,
- [Hana Hanzlikova](#) ORCID: [orcid.org/0000-0001-7235-7269](https://orcid.org/0000-0001-7235-7269)<sup>6,7</sup>,
- [Solji Park](#)<sup>8,9</sup>,
- [Seol Kyoung Jung](#)<sup>8,9</sup>,
- [Peter J. McHugh](#) ORCID: [orcid.org/0000-0002-8679-4627](https://orcid.org/0000-0002-8679-4627)<sup>3</sup>,
- [Andres Canela](#)<sup>10</sup>,
- [Chongyi Chen](#)<sup>11</sup>,
- [Rafael Casellas](#)<sup>8,9</sup>,
- [Keith W. Caldecott](#) ORCID: [orcid.org/0000-0003-4255-9016](https://orcid.org/0000-0003-4255-9016)<sup>6,7</sup>,
- [Michael E. Ward](#) ORCID: [orcid.org/0000-0002-5296-8051](https://orcid.org/0000-0002-5296-8051)<sup>2</sup> &

- [André Nussenzweig](#)   [ORCID: orcid.org/0000-0002-8952-7268<sup>1</sup>](#)

[Nature](#) volume **593**, pages 440–444 (2021)[Cite this article](#)

- 16k Accesses
- 2 Citations
- 326 Altmetric
- [Metrics details](#)

## Subjects

- [Molecular neuroscience](#)
- [Single-strand DNA breaks](#)

## Abstract

Defects in DNA repair frequently lead to neurodevelopmental and neurodegenerative diseases, underscoring the particular importance of DNA repair in long-lived post-mitotic neurons<sup>1,2</sup>. The cellular genome is subjected to a constant barrage of endogenous DNA damage, but surprisingly little is known about the identity of the lesion(s) that accumulate in neurons and whether they accrue throughout the genome or at specific loci. Here we show that post-mitotic neurons accumulate unexpectedly high levels of DNA single-strand breaks (SSBs) at specific sites within the genome. Genome-wide mapping reveals that SSBs are located within enhancers at or near CpG dinucleotides and sites of DNA demethylation. These SSBs are repaired by PARP1 and XRCC1-dependent mechanisms. Notably, deficiencies in XRCC1-dependent short-patch repair increase DNA repair synthesis at neuronal enhancers, whereas defects in long-patch repair reduce synthesis. The high levels of SSB repair in neuronal enhancers are therefore likely to be sustained by both short-patch and long-patch processes. These data provide the first evidence of site- and cell-type-specific SSB repair, revealing unexpected levels of localized and continuous DNA breakage in neurons. In addition, they suggest an

explanation for the neurodegenerative phenotypes that occur in patients with defective SSB repair.

## Access options

Subscribe to Journal

Get full journal access for 1 year

\$199.00

only \$3.90 per issue

[Subscribe](#)

All prices are NET prices.

VAT will be added later in the checkout.

Tax calculation will be finalised during checkout.

Rent or Buy article

Get time limited or full article access on ReadCube.

from \$8.99

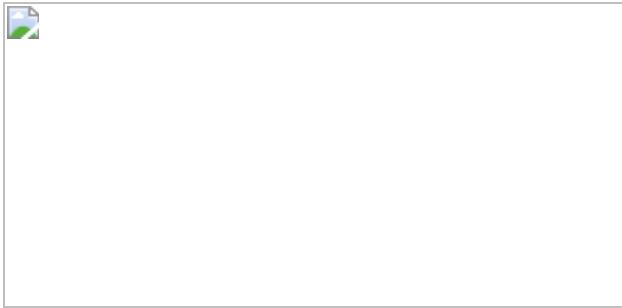
[Rent or Buy](#)

All prices are NET prices.

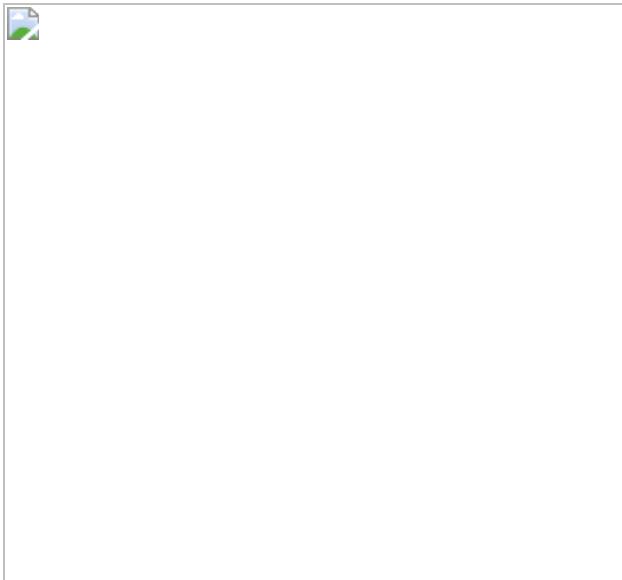
### Additional access options:

- [Log in](#)
- [Access through your institution](#)
- [Learn about institutional subscriptions](#)

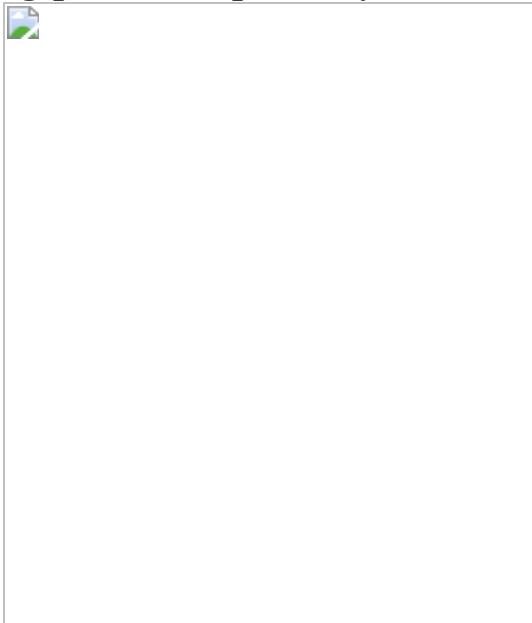
**Fig. 1: SAR-seq peaks occur within enhancers and are associated with PARP activation.**



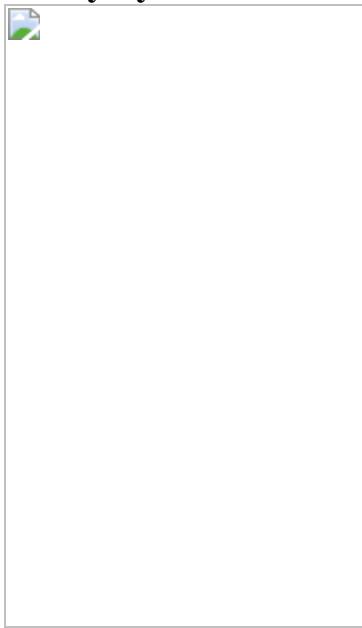
**Fig. 2: Single-strand breaks detected by S1 END-seq after incubation with ddN.**



**Fig. 3: Localized SSB repair in neurons comprises short-patch and long-patch sub-pathways.**



**Fig. 4: Localized SSB repair in neurons correlates with sites of oxidized 5-methylcytosine.**



## Data availability

The SAR-seq, S1 END-seq, Hi-C, RNA-seq, Seal and ChIP-seq data have been deposited in the Gene Expression Omnibus (GEO) database under accession number [GSE167259](#). iPS cell enhancers were identified using H3K4me1 ChIP-seq from an ENCODE iPS cell line (GSM2527632). Experimentally validated enhancers were from VISTA genome browser for enhancers: <https://enhancer.lbl.gov/>.

## References

1. 1.

Caldecott, K. W. Single-strand break repair and genetic disease. *Nat. Rev. Genet.* **9**, 619–631 (2008).

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

2. 2.

McKinnon, P. J. Genome integrity and disease prevention in the nervous system. *Genes Dev.* **31**, 1180–1194 (2017).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

3. 3.

Tubbs, A. & Nussenzweig, A. Endogenous DNA damage as a source of genomic instability in cancer. *Cell* **168**, 644–656 (2017).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

4. 4.

Miller, M. R. & Chinault, D. N. The roles of DNA polymerases alpha, beta, and gamma in DNA repair synthesis induced in hamster and human cells by different DNA damaging agents. *J. Biol. Chem.* **257**, 10204–10209 (1982).

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

5. 5.

Fernandopulle, M. S. et al. Transcription factor-mediated differentiation of human iPSCs into neurons. *Curr. Protoc. Cell Biol.* **79**, e51 (2018).

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

6. 6.

Wang, C. et al. Scalable production of iPSC-derived human neurons to identify tau-lowering compounds by high-content screening. *Stem Cell Reports* **9**, 1221–1233 (2017).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

7. 7.

Macheret, M. & Halazonetis, T. D. Intragenic origins due to short G1 phases underlie oncogene-induced DNA replication stress. *Nature* **555**, 112–116 (2018).

[ADS](#) [CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

8. 8.

Tubbs, A. et al. Dual roles of poly(dA:dT) tracts in replication initiation and fork collapse. *Cell* **174**, 1127–1142.e19 (2018).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

9. 9.

van der Raadt, J., van Gestel, S. H. C., Nadif Kasri, N. & Albers, C. A. ONECUT transcription factors induce neuronal characteristics and remodel chromatin accessibility. *Nucleic Acids Res.* **47**, 5587–5602 (2019).

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

10. 10.

Song, M. et al. Mapping *cis*-regulatory chromatin contacts in neural cells links neuropsychiatric disorder risk variants to target genes. *Nat. Genet.* **51**, 1252–1262 (2019).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

11. 11.

Visel, A., Minovitsky, S., Dubchak, I. & Pennacchio, L. A. VISTA Enhancer Browser—a database of tissue-specific human enhancers. *Nucleic Acids Res.* **35**, D88–D92 (2007).

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

12. 12.

Gupte, R., Liu, Z. & Kraus, W. L. PARPs and ADP-ribosylation: recent advances linking molecular functions to biological outcomes. *Genes Dev.* **31**, 101–126 (2017).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

13. 13.

Hanzlikova, H. & Caldecott, K. W. Perspectives on PARPs in S phase. *Trends Genet.* **35**, 412–422 (2019).

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

14. 14.

Gibson, B. A., Conrad, L. B., Huang, D. & Kraus, W. L. Generation and characterization of recombinant antibody-like ADP-ribose binding proteins. *Biochemistry* **56**, 6305–6316 (2017).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

15. 15.

Madabhushi, R. et al. Activity-induced DNA breaks govern the expression of neuronal early-response genes. *Cell* **161**, 1592–1605 (2015).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

16. 16.

Suberbielle, E. et al. Physiologic brain activity causes DNA double-strand breaks in neurons, with exacerbation by amyloid- $\beta$ . *Nat. Neurosci.* **16**, 613–621 (2013).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

17. 17.

Canela, A. et al. Topoisomerase II-induced chromosome breakage and translocation is determined by chromosome architecture and transcriptional activity. *Mol. Cell* **75**, 252–266.e8 (2019).

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

18. 18.

Gómez-Herreros, F. et al. TDP2 suppresses chromosomal translocations induced by DNA topoisomerase II during gene transcription. *Nat. Commun.* **8**, 233 (2017).

[ADS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

19. 19.

Canela, A. et al. DNA breaks and end resection measured genome-wide by end sequencing. *Mol. Cell* **63**, 898–911 (2016).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

20. 20.

Caldecott, K. W. XRCC1 protein; form and function. *DNA Repair* **81**, 102664 (2019).

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

21. 21.

Hanzlikova, H., Gittens, W., Krejcikova, K., Zeng, Z. & Caldecott, K. W. Overlapping roles for PARP1 and PARP2 in the recruitment of endogenous XRCC1 and PNKP into oxidized chromatin. *Nucleic Acids Res.* **45**, 2546–2557 (2017).

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

22. 22.

Caldecott, K. W. DNA single-strand break repair. *Exp. Cell Res.* **329**, 2–8 (2014).

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

23. 23.

Caldecott, K. W. Mammalian DNA base excision repair: dancing in the moonlight. *DNA Repair* **93**, 102921 (2020).

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

24. 24.

Tian, R. et al. CRISPR interference-based platform for multimodal genetic screens in human iPSC-derived neurons. *Neuron* **104**, 239–255.e12 (2019).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

25. 25.

Beard, W. A., Horton, J. K., Prasad, R. & Wilson, S. H. Eukaryotic base excision repair: new approaches shine light on mechanism. *Annu. Rev. Biochem.* **88**, 137–162 (2019).

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

26. 26.

DiGiuseppe, J. A., Hunting, D. J. & Dresler, S. L. Aphidicolin-sensitive DNA repair synthesis in human fibroblasts damaged with bleomycin is distinct from UV-induced repair. *Carcinogenesis* **11**, 1021–1026 (1990).

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

27. 27.

Poetsch, A. R. The genomics of oxidative DNA damage, repair, and resulting mutagenesis. *Comput. Struct. Biotechnol. J.* **18**, 207–219 (2020).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

28. 28.

Bansal, K., Yoshida, H., Benoist, C. & Mathis, D. The transcriptional regulator Aire binds to and activates super-enhancers. *Nat. Immunol.* **18**, 263–273 (2017).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

29. 29.

Puc, J. et al. Ligand-dependent enhancer activation regulated by topoisomerase-I activity. *Cell* **160**, 367–380 (2015).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

30. 30.

Kalasova, I. et al. Pathological mutations in PNKP trigger defects in DNA single-strand break repair but not DNA double-strand break repair. *Nucleic Acids Res.* **48**, 6672–6684 (2020).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

31. 31.

Whitehouse, C. J. et al. XRCC1 stimulates human polynucleotide kinase activity at damaged DNA termini and accelerates DNA single-strand break repair. *Cell* **104**, 107–117 (2001).

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

32. 32.

Lio, C. J. et al. TET methylcytosine oxidases: new insights from a decade of research. *J. Biosci.* **45**, 21 (2020).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

33. 33.

Kriaucionis, S. & Heintz, N. The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain. *Science* **324**, 929–930 (2009).

[ADS](#) [CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

34. 34.

Steinacher, R. et al. SUMOylation coordinates BERosome assembly in active DNA demethylation during cell differentiation. *EMBO J.* **38**, e99242 (2019).

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

35. 35.

Song, C. X. et al. Genome-wide profiling of 5-formylcytosine reveals its roles in epigenetic priming. *Cell* **153**, 678–691 (2013).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

36. 36.

Szulwach, K. E. et al. 5-hmC-mediated epigenetic dynamics during postnatal neurodevelopment and aging. *Nat. Neurosci.* **14**, 1607–1616 (2011).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

37. 37.

Hoch, N. C. et al. XRCC1 mutation is associated with PARP1 hyperactivation and cerebellar ataxia. *Nature* **541**, 87–91 (2017).

[ADS](#) [CAS](#) [PubMed](#) [Google Scholar](#)

38. 38.

Lodato, M. A. et al. Aging and neurodegeneration are associated with increased mutations in single human neurons. *Science* **359**, 555–559 (2018).

[ADS](#) [CAS](#) [PubMed](#) [Google Scholar](#)

39. 39.

Ried, D. A. et al. Incorporation of a nucleoside analog maps genome repair sites in post-mitotic human neurons. *Science* (in the press) (2021).

40. 40.

Watanabe, S. et al. MyoD gene suppression by Oct4 is required for reprogramming in myoblasts to produce induced pluripotent stem cells. *Stem Cells* **29**, 505–516 (2011).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

41. 41.

Akiyama, T. et al. Efficient differentiation of human pluripotent stem cells into skeletal muscle cells by combining RNA-based MYOD1-expression and POU5F1-silencing. *Sci. Rep.* **8**, 1189 (2018).

[ADS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

42. 42.

Selvaraj, S. et al. Screening identifies small molecules that enhance the maturation of human pluripotent stem cell-derived myotubes. *eLife* **8**,

e47970 (2019).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

43. 43.

Pawlowski, M. et al. Inducible and deterministic forward programming of human pluripotent stem cells into neurons, skeletal myocytes, and oligodendrocytes. *Stem Cell Reports* **8**, 803–812 (2017).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

44. 44.

Gilbert, L. A. et al. Genome-scale CRISPR-mediated control of gene repression and activation. *Cell* **159**, 647–661 (2014).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

45. 45.

Farias, G. G., Britt, D. J. & Bonifacino, J. S. Imaging the polarized sorting of proteins from the Golgi complex in live neurons. *Methods Mol. Biol.* **1496**, 13–30 (2016).

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

46. 46.

Kirwan, P., Jura, M. & Merkle, F. T. Generation and characterization of functional human hypothalamic neurons. *Curr. Protoc. Neurosci.* **81**, 3.33.1–3.33.24 (2017).

[Google Scholar](#)

47. 47.

Wong, N., John, S., Nussenzweig, A. & Canela, A. END-seq: an unbiased, high-resolution, and genome-wide approach to map DNA

double-strand breaks and resection in human cells. *Methods Mol. Biol.* **2153**, 9–31 (2021).

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

48. 48.

Bredemeyer, A. L. et al. DNA double-strand breaks activate a multi-functional genetic program in developing lymphocytes. *Nature* **456**, 819–823 (2008).

[ADS](#) [CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

49. 49.

Santos, M. A. et al. DNA-damage-induced differentiation of leukaemic cells as an anti-cancer barrier. *Nature* **514**, 107–111 (2014).

[ADS](#) [CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

50. 50.

Gibson, D. G. et al. Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods* **6**, 343–345 (2009).

[CAS](#) [Google Scholar](#)

51. 51.

Noordermeer, S. M. et al. The shieldin complex mediates 53BP1-dependent DNA repair. *Nature* **560**, 117–121 (2018).

[ADS](#) [CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

52. 52.

Cui, X. L. et al. A human tissue map of 5-hydroxymethylcytosines exhibits tissue specificity through gene and enhancer modulation. *Nat. Commun.* **11**, 6161 (2020).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

53. 53.

Dai, Q. & He, C. Syntheses of 5-formyl- and 5-carboxyl-dC containing DNA oligos as potential oxidation products of 5-hydroxymethylcytosine in DNA. *Org. Lett.* **13**, 3446–3449 (2011).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

54. 54.

Rao, S. S. et al. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* **159**, 1665–1680 (2014).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

55. 55.

Buenrostro, J. D., Giresi, P. G., Zaba, L. C., Chang, H. Y. & Greenleaf, W. J. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat. Methods* **10**, 1213–1218 (2013).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

56. 56.

Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10**, R25 (2009).

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

57. 57.

Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

58. 58.

Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).

[CAS](#) [Google Scholar](#)

59. 59.

Li, H. et al. The sequence alignment/map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).

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

60. 60.

Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

61. 61.

Zhang, Y. et al. Model-based analysis of ChIP–Seq (MACS). *Genome Biol.* **9**, R137 (2008).

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

62. 62.

Zang, C. et al. A clustering approach for identification of enriched domains from histone modification ChIP-seq data. *Bioinformatics* **25**, 1952–1958 (2009).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

63. 63.

Amemiya, H. M., Kundaje, A. & Boyle, A. P. The ENCODE blacklist: identification of problematic regions of the genome. *Sci. Rep.* **9**, 9354 (2019).

[ADS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

64. 64.

Salmon-Divon, M., Dvinge, H., Tammoja, K. & Bertone, P. PeakAnalyzer: genome-wide annotation of chromatin binding and modification loci. *BMC Bioinformatics* **11**, 415 (2010).

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

65. 65.

Harrow, J. et al. GENCODE: the reference human genome annotation for The ENCODE Project. *Genome Res.* **22**, 1760–1774 (2012).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

66. 66.

Whyte, W. A. et al. Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell* **153**, 307–319 (2013).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

67. 67.

Huang, W., Sherman, B. T. & Lempicki, R. A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protocols* **4**, 44–57 (2009).

[CAS](#) [Google Scholar](#)

68. 68.

Machanick, P. & Bailey, T. L. MEME-ChIP: motif analysis of large DNA datasets. *Bioinformatics* **27**, 1696–1697 (2011).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

69. 69.

Durand, N. C. et al. Juicebox provides a visualization system for Hi-C contact maps with unlimited zoom. *Cell Syst.* **3**, 99–101 (2016).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

70. 70.

Kent, W. J., Zweig, A. S., Barber, G., Hinrichs, A. S. & Karolchik, D. BigWig and BigBed: enabling browsing of large distributed datasets. *Bioinformatics* **26**, 2204–2207 (2010).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

71. 71.

Kent, W. J. et al. The human genome browser at UCSC. *Genome Res.* **12**, 996–1006 (2002).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

72. 72.

Ramírez, F. et al. deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res.* **44**, W160–W165 (2016).

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

[Download references](#)

## Acknowledgements

We thank A. Rao, S. Wilson, S. Ruiz, N. Lopez-Bigas, D. Wilson III, Y. Sun, Y. Pommier, M. Adamowicz, P. Meltzer, and K. Ge for discussions and reagents; R. Chari for constructing the CAS9D10A inducible nickase; and the CCR genomics core for help with sequencing. K.W.C. is supported by Programme Grants from the UK Medical Research Council (MR/P010121/1), Cancer Research-UK (C6563/A7322), is an ERC Advanced Investigator (SIDSCA 694996) and received a Royal Society Wolfson Research Merit Award. The M.E.W. laboratory is supported by the NINDS Intramural Research Program, the Chan Zuckerberg Initiative, and the Packard ALS Center. S.E.H. received funding from the BrightFocus Foundation. The A.N. laboratory is supported by the Intramural Research Program of the NIH, an Ellison Medical Foundation Senior Scholar in Aging Award (AG-SS-2633-11), Department of Defense Awards (W81XWH-16-1-599 and W81XWH-19-1-0652), an Alex's Lemonade Stand Foundation Award, and an NIH Intramural FLEX Award.

## Author information

### Author notes

1. These authors contributed equally: Wei Wu, Sarah E. Hill, William J. Nathan

## Affiliations

1. Laboratory of Genome Integrity, National Cancer Institute, NIH, Bethesda, MD, USA

Wei Wu, William J. Nathan, Jacob Paiano, Elsa Callen, Dongpeng Wang, Kenta Shinoda, Niek van Wietmarschen, Dali Zong, Raphael Pavani & André Nussenzweig

2. National Institute of Neurological Disorders and Stroke, NIH, Bethesda, MD, USA

Sarah E. Hill, Jennifer M. Colón-Mercado, Maia Parsadanian & Michael E. Ward

3. Department of Oncology, MRC Weatherall Institute of Molecular Medicine, University of Oxford, John Radcliffe Hospital, Oxford, UK

William J. Nathan & Peter J. McHugh

4. Eunice Kennedy Shriver National Institute of Child Health and Human Development, Bethesda, MD, USA

Raffaella De Pace & Steve Coon

5. National Eye Institute, NIH, Bethesda, MD, USA

Han-Yu Shih

6. Department of Genome Dynamics, Institute of Molecular Genetics of the Czech Academy of Sciences, Prague, Czech Republic

Hana Hanzlikova & Keith W. Caldecott

7. Genome Damage and Stability Centre, School of Life Sciences, University of Sussex, Brighton, UK

Hana Hanzlikova & Keith W. Caldecott

8. Lymphocyte Nuclear Biology, National Institute of Arthritis and Musculoskeletal and Skin Diseases and National Cancer Institute, NIH, Bethesda, MD, USA

Solji Park, Seol Kyoung Jung & Rafael Casellas

9. NIH Regulome Project, NIH, Bethesda, MD, USA

Solji Park, Seol Kyoung Jung & Rafael Casellas

10. The Hakubi Center for Advanced Research and Radiation Biology Center, Graduate School of Biostudies, Kyoto University, Kyoto, Japan

Andres Canela

11. Laboratory of Biochemistry and Molecular Biology, National Cancer Institute, NIH, Bethesda, MD, USA

Chongyi Chen

## Authors

1. Wei Wu

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

2. Sarah E. Hill

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

3. William J. Nathan

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

4. Jacob Paiano

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

5. Elsa Callen

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

6. Dongpeng Wang

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

7. Kenta Shinoda

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

8. Niek van Wietmarschen  
[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

9. Jennifer M. Colón-Mercado  
[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

10. Dali Zong  
[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

11. Raffaella De Pace  
[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

12. Han-Yu Shih  
[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

13. Steve Coon  
[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

14. Maia Parsadanian  
[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

15. Raphael Pavani  
[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

16. Hana Hanzlikova

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

17. Solji Park

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

18. Seol Kyoung Jung

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

19. Peter J. McHugh

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

20. Andres Canela

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

21. Chongyi Chen

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

22. Rafael Casellas

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

23. Keith W. Caldecott

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

24. Michael E. Ward

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

25. André Nussenzweig

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

## Contributions

W.W., S.E.H., W.J.N., J.P., K.W.C., M.E.W. and A.N. conceived, designed and analysed the experiments; W.W. designed bioinformatics pipelines, performed data analysis and designed the figures; S.E.H., E.C. and M.P. prepared i<sup>3</sup>Neurons for experiments; S.E.H. and E.C. performed gene knockdown, i<sup>3</sup>Neuron treatments, dissociation and collections in i<sup>3</sup>Neurons; J.P. and D.W. performed SAR-seq; W.J.N., J.P. and D.W. performed ChIP-seq; W.J.N. and N.V.W. performed S1 END-seq. S.E.H., E.C. and W.J.N. performed immunofluorescence imaging; A.C. developed SAR-seq in the A.N. laboratory; W.J.N. developed S1 END-seq; S.E.H. and M.E.W. selected guides for gene knockdown; K.S. performed ATAC-seq, did cloning for knock-down experiments and qRT–PCR; D.W. performed SEAL experiments; J.C.-M. prepared iMuscle cells; R.D.P. prepared primary rat neurons; S.E.H. treated and collected iMuscle and primary rat neurons; H.-Y.S. and S.C. conducted sequencing experiments in i<sup>3</sup>Neurons that were informative to the study; R.P. made MCF10A cells with Cas9-nickase; D.Z. and C.C. assisted with experiments; S.P. performed Hi-C, and S.K.J. and R.C. analysed the Hi-C data; H.H., P.J.M. and C.C. provided insights; K.W.C., M.E.W. and A.N. wrote the paper with input from all co-authors; M.E.W. and A.N. supervised the study.

## Corresponding authors

Correspondence to [Keith W. Caldecott](#) or [Michael E. Ward](#) or [André Nussenzweig](#).

## Ethics declarations

### Competing interests

The authors declare no competing interests.

### Additional information

**Peer review information** *Nature* thanks Michael Weinfeld and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

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

### Extended data figures and tables

#### [Extended Data Fig. 1 Discrete genomic loci in neurons are associated with ongoing DNA synthesis.](#)

**a**, Schematic of SAR-seq (DNA synthesis associated with repair sequencing) methodology. Neurons grown in culture (1) are incubated with EdU to label sites of DNA repair synthesis (2). The incorporated genomic EdU is then conjugated to biotin via click chemistry (3), sheared by sonication to fragments of 150–200 bp and captured with streptavidin beads (4). Enriched DNA sequences are then PCR-amplified and subjected to next-generation sequencing (5). **b**, Genome browser screenshot displaying SAR-seq profiles as normalized read density (RPM) for i<sup>3</sup>Neurons. Three independent biological replicates are shown as well as input. Neurons were labelled with EdU for 18 h and collected on day 7 after induction of neuronal differentiation. All coordinates provided are from the hg19 reference genome for i<sup>3</sup>Neurons. **c**, Venn diagram showing the overlap of

SAR-seq peaks in i<sup>3</sup>Neurons for three independent biological replicates. **d**, Scatter plots showing correlations of SAR-seq intensities (RPKM) between three replicates in i<sup>3</sup>Neurons. Pearson correlation coefficients and *P* values are indicated. **e**, Left, representative images of EdU–biotin staining (green) showing cell proliferation in iPS cells, but not in post-mitotic i<sup>3</sup>Neurons. i<sup>3</sup>Neurons were treated with EdU on day 3 or day 6 and fixed on day 7. iPS cells were treated with EdU for 24 h and fixed. Cells were counterstained with DAPI (blue). Different imaging conditions were used for iPS cells and i<sup>3</sup>Neurons in the representative images. Right, quantification of EdU-positive cells. Each dot represents the percentage of dividing cells in one image (iPS cell:  $n = 8$  images,  $n = 397/410$  cells EdU-positive; i<sup>3</sup>Neuron day 3:  $n = 8$  images,  $n = 35/483$  cells EdU-positive; i<sup>3</sup>Neuron day 7:  $n = 8$  images,  $n = 0/523$  cells EdU-positive). Data are mean ± s.e.m. and are representative of three independent experiments.

## Extended Data Fig. 2 Genomic characteristics of SAR-seq peaks.

**a**, Genome browser screenshot showing SAR-seq in i<sup>3</sup>Neurons treated with hydroxyurea (HU,  $n = 1$ ) or polymerase- $\alpha$  inhibitor (POL $\alpha$ i,  $n = 2$ ). NT, non-treated. **b**, Scatter plots showing SAR-seq intensities (RPKM) for HU-(left) and POL $\alpha$ i- (right) treated compared to non-treated (NT) samples. **c**, Flow cytometry cell cycle profile of pre-B cells treated or non-treated (NT) with POL $\alpha$ i. Cells were pulsed with EdU for 30 min before being collected for flow cytometry. Cells were counterstained with DAPI ( $n = 1$ ). For an example of the gating strategy used for flow cytometry in **c** and in Extended Data Fig. 9b, see Supplementary Fig. 1. **d**, Histogram of individual SAR-seq peak widths, revealing a mean peak width of 901 bp. **e**, Genome browser screenshot showing SAR-seq in i<sup>3</sup>Neurons collected after 1, 2, 4, 8, or 18 h of EdU incubation ( $n = 2$  for each). **f**, Fraction of EdU labelling in i<sup>3</sup>Neurons (relative to maximum labelling at 18 h) as a function of time, calculated from the top 2,000 peaks. Red points and error bars represent mean and s.d. of the relative levels of EdU measured from experimental data. Black line represents the theoretical model after fitting, with *k* being the rate of EdU labelling.

## Extended Data Fig. 3 SAR-seq enrichment at neuronal intragenic regions.

**a**, Genome browser screenshot of chromosome 7 showing lack of localized DNA synthesis in two independent biological replicates of SAR-seq performed in iMuscle cells ( $n = 2$ ) incubated with EdU for 18 h compared to SAR-seq in i<sup>3</sup>Neurons as well as input DNA. **b**, Genome browser screenshot displaying SAR-seq peak at a representative AsiSI restriction enzyme site (tick mark). AsiSI expression was induced for 18 h (+Dox,  $n = 1$ ) compared with non-treated (−Dox,  $n = 1$ ) conditions in G0-arrested, Abelson virus-transformed mouse pre-B cells as described<sup>19</sup>. **c**, Distribution of SAR-seq peaks with respect to different genomic features compared to genome-wide distribution of the hg19 human reference genome. Promoters are defined as 1 kb upstream of transcription start sites. Distal intergenic represents promoter-excluded intergenic regions. **d**, Fold enrichment of SAR-seq peaks in intergenic regions, intragenic regions and expressed genes compared to 1,000 sets of randomly shuffled regions of the same sizes and chromosome distribution (one-sided Fisher's exact test, \*\*\*\* $P < 2.2 \times 10^{-16}$ ; NS, not significant). **e**, Left, scatterplot showing correlation of SAR-seq intensity (RPKM) with transcript level of genes containing SAR-seq peaks measured by RNA-seq (FPKM;  $n = 3$ ) in i<sup>3</sup>Neurons. Seventy-one per cent of SAR-seq peaks are at expressed genes ( $FPKM \geq 0.1$ ; red dashed line,  $FPKM = 0.1$ ). Right, correlation of SAR-seq intensity with transcript levels of linked genes determined by pcHiC in i<sup>3</sup>Neurons (red dashed line,  $FPKM = 0.1$ ). Spearman correlation coefficients and  $P$  values are indicated. **f**, Genome browser screenshot comparing SAR-seq with strand-specific SAR-seq ( $n = 3$ ), which discriminates which strand is labelled with EdU in i<sup>3</sup>Neurons. Both strands show labelling in three biological replicates. Strand-specific SAR-seq reads are separated into positive (black) and negative (grey) strands.

## Extended Data Fig. 4 Motif discovery of SAR peaks and comparison with ATAC-seq peaks.

**a**, Motif analysis for sequences within 500 bp of the summit of the top 5,000 SAR-seq peaks in i<sup>3</sup>Neurons. Top, the best motif discovered by the

MEME suite (1,384 out of 5,000 sites have this motif). Middle, TOMTOM motif tool used to compare SAR-seq motif shown above with databases of known motifs. The transcription factor ONECUT1 was identified as the most similar motif and its consensus sequence is shown. *P* value for motif comparison with ONECUT1 motif is indicated. Bottom, position distribution of the best motif (top) within 500 bp of the SAR-seq peak summit. The best motif is centred on the SAR-seq peak summit. **b**, Venn diagram illustrating the overlap between H3K4me1 ChIP-seq, ATAC-seq and SAR-seq peaks in i<sup>3</sup>Neurons. The statistical significance of the overlaps between SAR-seq, H3K4me1 ChIP-seq and ATAC-seq peaks was determined using randomly shuffled data sets (*n* = 1,000) by one-sided Fisher's exact test (the *P* value for overlap between H3K4me1 ChIP-seq and SAR-seq peaks is  $P < 2.2 \times 10^{-16}$ , and for ATAC-seq/SAR-seq peaks is  $P < 2.2 \times 10^{-16}$ ). Fraction of different overlapping groups are labelled in red for SAR-seq peaks, green for ATAC-seq peaks and blue for H3K4me1 ChIP-seq peaks. **c**, Heat maps of SAR-seq and ATAC-seq signals within 1 kb of SAR-seq peak summits in i<sup>3</sup>Neurons, ordered by SAR-seq intensity. **d**, Scatter plot comparing widths of ATAC-seq peaks and SAR-seq peaks for the top 10,000 overlapping peaks in i<sup>3</sup>Neurons. Pearson correlation coefficient and *P* value are indicated. **e**, Distribution of SAR-seq and ATAC-seq peaks with respect to different genomic features compared to genome-wide distribution on the hg19 human reference genome. Promoters are defined as 1 kb upstream of transcription start sites and distal intergenic represents promoter-excluded intergenic regions. **f**, Fold enrichment of SAR-seq and ATAC-seq peaks located at enhancers (black) and promoters (grey) compared to 1,000 sets of randomly shuffled regions of the same sizes and chromosome distributions, respectively (one-sided Fisher's exact test, \*\*\* $P < 2.2 \times 10^{-16}$ , NS:  $P = 0.0783$ , not significant). **g**, Heat map of RNA Pol II (*n* = 1) ChIP-seq and SAR-seq in i<sup>3</sup>Neurons for 1 kb on either side of the transcription start site (TSS) in i<sup>3</sup>Neurons, ordered by Pol II ChIP-seq intensity.

## Extended Data Fig. 5 The correlation between SAR-seq and chromatin features.

**a**, Heat maps of SAR-seq and ChIP-seq data for enhancer markers (H3K4me1, H3K27ac and MLL4), other chromatin markers at accessible regions (H3K4me3, H3K36me3 and CTCF) and chromatin silencing markers (H3K27me3 and H3K9me3) for 1 kb on either side of the SAR-seq peak summit in i<sup>3</sup>Neurons, ordered by SAR-seq intensity. **b**, Scatter plots showing the correlation between SAR-seq intensity and ATAC-seq, H3K4me1, H3K27ac and MLL4 ChIP-seq intensities (RPKM) for 1 kb on either side of the SAR-seq peak summits in i<sup>3</sup>Neurons. Spearman correlation coefficients and *P* values are indicated. **c**, Fold enrichment of SAR-seq peaks at ATAC-seq peaks, ChIP-seq peaks of enhancer-related marks (H3K4me1, H3K27ac and MLL4), additional chromatin marks at accessible regions (H3K4me3, H3K36me3 and CTCF) (all *n* = 1) and chromatin silencing marks (H3K27me3 and H3K9me3) (both *n* = 1) in i<sup>3</sup>Neurons. *n* = 1,000 randomly shuffled data sets were generated to test the significance using one-sided Fisher's exact test (*P* < 2.2 × 10<sup>-16</sup> for H3K4me1, H3K27ac, ATAC-seq and MLL4; *P* = 1.85 × 10<sup>-316</sup> for H3K4me3; *P* = 0.00116 for CTCF; \*\**P* < 0.001, \*\*\*\**P* < 0.00001, NS, not significant). **d**, Heat maps of SAR-seq and H3K4me1 ChIP-seq signal for 1 kb on either side of the SAR-seq peak summit in primary rat cortical neurons, ordered by SAR-seq intensity. **e**, Venn diagram showing the overlap between H3K4me1 and SAR-seq peaks in rat primary neurons. *n* = 1,000 randomly shuffled data sets were generated to test the significance using one-sided Fisher's exact test: *P* < 2.2 × 10<sup>-16</sup>. **f**, Scatter plot showing the correlation between SAR-seq and H3K4me1 ChIP-seq intensities (RPKM) for 1 kb on either side of the SAR-seq peak summits in rat primary neurons. Spearman correlation coefficient and *P* values are indicated.

## Extended Data Fig. 6 SAR-seq enrichment at neuronal enhancers.

**a**, Box plot showing Hi-C contacts (left, *n* = 4) and H3K27ac (right, *n* = 1) levels at enhancers with (red) or without (grey) SAR-seq peaks. Contacts were defined as Hi-C interactions between an H3K27ac<sup>+</sup> enhancer (with and without SAR) and its closest promoter within the TAD domain. For comparative purposes, observed contacts were normalized to expected

contacts. Centre line, median; box limits, 25th and 75th percentiles; whisker indicates  $1.5\times$  interquartile range. Statistical significance was determined using the one-sided Wilcoxon rank-sum test.  $P = 1.64 \times 10^{-225}$  for HiC contacts and  $P < 2.2 \times 10^{-16}$  for H3K27ac intensity ( $****P < 0.00001$ ). **b**, Fold enrichment of SAR-seq peaks at in vivo-validated enhancers from the VISTA Enhancer Browser database that overlap with H3K4me1 ChIP-seq peaks (left,  $P = 1.42 \times 10^{-53}$ ) or at promoter-interacting regions determined by pcHiC (right,  $P = 1.574 \times 10^{-09}$ ).  $n = 1,000$  randomly shuffled datasets were generated to test significance using one-sided Fisher's exact test ( $**P < 0.001$ ,  $****P < 0.00001$ ). **c**, **d**, Genome browser screenshots showing SAR-seq, H3K4me1 and H3K27ac ChIP-seq, and pcHiC and Hi-C profiles at representative enhancers (highlighted in orange) interacting with the *CDK5RAP3* promoter (**c**) and the *DRD2* promoter (**d**). Both enhancers have been validated to promote transcription of their respective genes using CRISPR techniques in i<sup>3</sup>Neurons<sup>10</sup>. The *CDK5RAP3* enhancer also overlaps with in vivo-validated enhancers from the VISTA Enhancer Browser database. In the Hi-C contact matrix (**c**, bottom) the intensity of each pixel represents the normalized number of contacts between a pair of loci. The maximum intensity is indicated at the lower left corner. **e**, Pie chart showing distribution of i<sup>3</sup>Neuron SAR-seq peaks in iPS cell-specific, i<sup>3</sup>Neuron-specific and shared iPS cell and i<sup>3</sup>Neuron enhancers. Approximately 56 million and 49 million single end reads were sequenced for the H3K4me1 ChIP-seq in iPS cell and i<sup>3</sup>Neurons, respectively, with approximately 100,000 peaks called in both cell types. **f**, Top biological processes enriched for genes containing the 2,000 most intense SAR-seq peaks determined by GO analysis. The x-axis represents the enrichment value as the logarithm of FDR. **G**, H3K27ac signal at enhancers in i<sup>3</sup>Neurons ranked by H3K27ac ChIP-seq intensity. Red dashed line indicates the inflection point of the H3K27ac signal used to identify super-enhancers (cutoff 1,000). Accordingly, 1,385 enhancers were defined as super-enhancers. **h**, Bar graph showing the fraction of super-enhancers (left) and conventional enhancers (right) that overlap with SAR-seq peaks. The super-enhancers in the i<sup>3</sup>Neurons were defined by H3K27ac ChIP-seq intensity in **g**.

## Extended Data Fig. 7 PARP and XRCC1 recruited to DNA repair sites.

**a**, Quantification of PAR levels with ( $n = 422$  cells) and without ( $n = 541$  cells) MMS treatment (red line, mean). Each dot represents one cell. Statistical significance was determined using two sided Mann–Whitney test ( $****P < 0.0001$ ). Data are representative of three independent experiments. **b**, Anti-ADP-ribose ChIP–seq signal ( $n = 1$ ) at an AsiSI restriction enzyme cut site (tick mark) in Abelson virus-transformed mouse pre-B cells. Cells were arrested in G0, and AsiSI double-strand breaks were induced for 18 h before ChIP. ADP-ribose is enriched at cleaved AsiSI sites and is increased by 20 min treatment with PARGi before fixation (AsiSI + PARGi), which is indicative of the presence of PAR. **c**, Genome browser screenshot illustrating the overlap between SAR-seq ( $n = 3$ ), PAR ( $n = 1$ ) and XRCC1 ( $n = 1$ ) ChIP–seq signals in i<sup>3</sup>Neurons. Cells for PAR ChIP–seq were incubated with PARGi for 20 min before fixation.

## Extended Data Fig. 8 Mapping regions of DNA damage and repair in neurons.

**a**, Genome browser example of SAR-seq profiles in non-treated (NT) or etoposide (ETO)-treated (18 h, 50 μM) i<sup>3</sup>Neurons. Data are from two biological replicates. **b**, Heat maps for SAR-seq in nontreated (NT) or etoposide- (ETO) treated (18 h, 50 μM) i<sup>3</sup>Neurons at –2 kb to +5 kb from the transcription start sites (TSS), ordered by ETO SAR-seq intensity. **c**, Immunofluorescence staining of the DSB markers γH2AX (red) and 53BP1 (green) in non-treated or ETO-treated (1 h) i<sup>3</sup>Neurons. Data are representative of three independent experiments. **d**, Genome browser showing SAR-seq and END-seq profiles in non-treated i<sup>3</sup>Neurons. END-seq, which detects DSBs specifically<sup>19</sup>, does not detect any enriched signal (that is, above background) at SAR-seq peaks. END-seq signals are separated into positive (black) and negative (grey) strands. END-seq data are representative of two independent experiments. **e**, Heat maps of SAR-seq and XRCC1 ChIP–seq ( $n = 1$ ) for 1 kb on either side of SAR-seq peak summits in cultured rat primary neurons, ordered by SAR-seq intensity. **f, g**, Scatter plots showing the correlation between SAR-seq and XRCC1 ChIP–

seq intensities (RPKM) for 1 kb before and after SAR-seq peak summits in i<sup>3</sup>Neurons (**f**) and rat primary neurons (**g**). Spearman correlation coefficients and *P* values are indicated.

### **Extended Data Fig. 9 S1 END-seq mapping of SSBs.**

**a**, Top, genome browser screenshot showing profiles of SAR-seq and S1 END-seq. Agarose plugs were incubated with or without the restriction enzyme Nt.BspQI before S1 treatment (*n* = 1). Bottom, expanded views of Nt.BspQI sites (tick mark) show detection of S1 END-seq upon Nt.BspQI treatment. S1 END-seq reads are separated into positive (black) and negative (grey) strands. **b**, Flow cytometry profile of G1-arrested MCF10A cells pulsed with EdU. MCF10A cells were treated with palbociclib (1  $\mu$ M) for 48 h to arrest cells in G1. Doxycycline was added in the last 24 h to induce nickase expression. Data are representative of three independent experiments. For an example of the gating strategy used for flow cytometry in **b** and Extended Data Fig. [2c](#), see Supplementary Fig. [1](#). **c**, Genome browser screenshot showing S1 END-seq profiles at three Cas9 nickase targeting sites (tick marks: sgRNAs 1–3) in G1-arrested MCF10A cells treated with doxycycline (+Dox) to induce Cas9 expression (*n* = 1). S1 END-seq signals are separated into positive (black) and negative (grey) strands. **d**, Heat maps of SAR-seq and S1 END-seq signals in i<sup>3</sup>Neurons with ddN or without incubation (NT) for 1 kb on either side of the SAR-seq peak summits, ordered by SAR-seq intensity. **e**, Venn diagram showing the overlap between S1 END-seq peaks incubated with ddN and SAR-seq peaks in i<sup>3</sup>Neurons. *n* = 1,000 randomly shuffled datasets were generated to test the significance using one-sided Fisher's exact test: *P* < 2.2  $\times$  10<sup>-16</sup>. **f**, Distribution of the size of the gaps between positive- and negative-strand S1 END-seq peak summits in i<sup>3</sup>Neurons incubated with ddN. The median gap size is 25 bp (red dashed line). The positive-strand peak represents the right end and the negative-strand peak represents the left end of a detected DSB. **g**, Aggregate plot showing the distribution of CG dinucleotides (black) for 1 kb on either side of SAR-seq peak summits overlaid with SAR-seq signal (red). **h**, Aggregate plot showing the distribution of CG dinucleotides (black) for 1 kb on either side of SAR-seq peak summits or summits of ATAC-seq peaks (green) that are H3K4me1 positive but do not overlap with SAR-seq.

## Extended Data Fig. 10 PARP, XRCC1 or POLB deficiency increases SAR.

**a**, Quantitative RT–PCR analysis showing the *PARP1* mRNA transcript level in i<sup>3</sup>Neurons after CRISPRi knockdown (sgControl, control non-targeting sgRNA; sg*PARP1*, an sgRNA targeting *PARP1*), cultured in parallel with samples used for SAR-seq.  $P = 7.72 \times 10^{-7}$  by unpaired two-tailed Student's *t*-test; \*\*\* $P < 0.00001$  ( $n = 3$ ). **b**, Genome browser screenshot displaying SAR-seq profiles from two biological replicates of i<sup>3</sup>Neurons treated with the PARP inhibitors olaparib, talazoparib, or veliparib, or using CRISPRi-mediated knockdown with sgControl or sg*PARP1*, in duplicates. NT, not treated. **c**, Quantitative RT–PCR analysis showing *XRCC1* mRNA transcript level in i<sup>3</sup>Neurons after CRISPRi knockdown, cultured in parallel with samples used for SAR-seq.  $P = 1.88 \times 10^{-5}$  by unpaired two-tailed Student's *t*-test; \*\*\* $P < 0.00001$  ( $n = 3$ ). **d**, Genome browser screenshots of SAR-seq profiles in i<sup>3</sup>Neurons expressing sgControl or sg*XRCC1*, in duplicate. **e**, Venn diagram showing the overlap of SAR-seq peaks between i<sup>3</sup>Neurons expressing sgControl and those expressing sg*XRCC1*.  $n = 1,000$  random datasets were generated to test the significance of overlap (one-sided Fisher's exact test:  $P < 2.2 \times 10^{-16}$ ). **f**, Quantitative RT–PCR analysis showing *POLB* mRNA transcript levels in i<sup>3</sup>Neurons after CRISPRi knockdown, cultured in parallel with samples used for SAR-seq.  $P = 6.98 \times 10^{-5}$  for sg*POLB1* and  $5.82 \times 10^{-5}$  for sg*POLB2* by unpaired two-tailed Student's *t*-test; \*\*\* $P < 0.00001$  ( $n = 3$ ). **g**, Western blot showing POLB protein levels in i<sup>3</sup>Neurons after CRISPRi knockdown, cultured in parallel with samples used for SAR-seq ( $n = 1$ ). For gel source data, see Supplementary Fig. 2. **h**, Genome browser screenshots of SAR-seq profiles from two biological replicates of i<sup>3</sup>Neurons expressing sgControl or sg*POLB*, in duplicate. **i**, Genome browser screenshots of SAR-seq profiles from two biological replicates of i<sup>3</sup>Neurons expressing sgControl or sg*POLB*. Cells were pre-treated or not treated (NT) with 50  $\mu$ M Aph for 24 h, and then also during incubation with EdU.

## Extended Data Fig. 11 Localized SSB repair in neurons correlates with sites of oxidized 5-methylcytosine.

**a**, Genome browser screenshot of chromosome 7 showing SAR-seq profiles from two biological replicates and input DNA in i<sup>3</sup>Neurons without MMS treatment (NT,  $n = 2$ ) or after treatment with 0.1 mg ml<sup>-1</sup> MMS ( $n = 2$ ) for the final 15 min of an 18-h incubation with EdU. After streptavidin pull-down and PCR amplification, total DNA was quantified (NT rep 1: 0.95 µg; NT rep 2: 1.7 µg; MMS rep 1: 3.8 µg; MMS rep 2: 4.5 µg). Stochastic DNA damage results in loss of DNA synthesis at recurrent sites. **b**, Quantitative RT-PCR analysis showing *PNKP* mRNA transcript level in i<sup>3</sup>Neurons after CRISPRi knockdown, cultured in parallel with samples used for SAR-seq.  $P = 0.00015$  by unpaired two-tailed Student's *t*-test; \*\*\* $P < 0.0001$  ( $n = 3$ ). **c**, Bottom, heat maps of SAR-seq intensities for 1 kb on either side of SAR-seq peak summits for i<sup>3</sup>Neurons expressing sgControl ( $n = 2$ ) or sg*PNKP* ( $n = 2$ ). Top, aggregate plots of SAR-seq intensity. **d**, Bottom, heat map of SAR-seq intensity for 1 kb on either side of the transcription start site in i<sup>3</sup>Neurons, ordered by SAR-seq intensity. i<sup>3</sup>Neurons expressing sgControl or sg*PNKP* were either not treated (NT,  $n = 2$ ) or treated with 25 µM camptothecin (CPT,  $n = 2$ ) during incubation with EdU. Top, aggregate plots of SAR-seq intensity. **e**, Scatter plots showing correlations of intensities (RPKM) between SSBs (ddN S1 END-seq) and 5fC or 5hmC, respectively, for 1 kb on either side of SAR-seq peak summits for i<sup>3</sup>Neurons. Spearman correlation coefficients and  $P$  values are indicated. **f**, Scatter plots showing correlations of intensities (RPKM) between SAR-seq for i<sup>3</sup>Neurons expressing sg*POLB* and ddN S1 END-seq, 5fC, or 5hmC for 1 kb on either side of SAR-seq peak summits. Spearman correlation coefficients and  $P$  values are indicated.

## Supplementary information

### Supplementary Information

This file contains Supplementary Figure 1 (Flow cytometry gating) and Supplementary Figure 2 (Source Data gels).

## Reporting Summary

## Rights and permissions

### Reprints and Permissions

## About this article



Check for  
updates

## Cite this article

Wu, W., Hill, S.E., Nathan, W.J. *et al.* Neuronal enhancers are hotspots for DNA single-strand break repair. *Nature* **593**, 440–444 (2021).  
<https://doi.org/10.1038/s41586-021-03468-5>

### Download citation

- Received: 28 November 2020
- Accepted: 17 March 2021
- Published: 25 March 2021
- Issue Date: 20 May 2021
- DOI: <https://doi.org/10.1038/s41586-021-03468-5>

## Further reading

- [Brain DNA damage hotspots](#)
  - Kim Baumann

*Nature Reviews Molecular Cell Biology* (2021)

- **DNA repair in neurons and its possible link to the epigenetic machinery at enhancers**

◦ Gerd P Pfeifer

*Epigenomics* (2021)

## Comments

By submitting a comment you agree to abide by our [Terms](#) and [Community Guidelines](#). If you find something abusive or that does not comply with our terms or guidelines please flag it as inappropriate.

[Access through your institution](#)

[Change institution](#)

[Buy or subscribe](#)

Advertisement

---

This article was downloaded by **calibre** from <https://www.nature.com/articles/s41586-021-03468-5>

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

Structure and dynamics of a mycobacterial type VII secretion system

[Download PDF](#)

- Article
- Open Access
- [Published: 12 May 2021](#)

# Structure and dynamics of a mycobacterial type VII secretion system

- [Catalin M. Bunduc](#) ORCID: [orcid.org/0000-0001-6633-1935](#)<sup>1,2,3,4</sup>,
- [Dirk Fahrenkamp](#) ORCID: [orcid.org/0000-0003-1149-5910](#)<sup>1,2,3</sup>,
- [Jiri Wald](#) ORCID: [orcid.org/0000-0002-1539-6488](#)<sup>1,2,3</sup>,
- [Roy Ummels](#)<sup>5</sup>,
- [Wilbert Bitter](#) ORCID: [orcid.org/0000-0001-8347-6511](#)<sup>4,5</sup>,
- [Edith N. G. Houben](#) ORCID: [orcid.org/0000-0002-8805-2002](#)<sup>4</sup> &
- [Thomas C. Marlovits](#) ORCID: [orcid.org/0000-0001-8106-4038](#)<sup>1,2,3</sup>

*Nature* volume 593, pages 445–448 (2021) [Cite this article](#)

- 9308 Accesses
- 136 Altmetric
- [Metrics details](#)

## Subjects

- [Bacterial pathogenesis](#)
- [Bacterial secretion](#)
- [Bacterial structural biology](#)
- [Cryoelectron microscopy](#)
- [Membrane proteins](#)

## Abstract

*Mycobacterium tuberculosis* is the cause of one of the most important infectious diseases in humans, which leads to 1.4 million deaths every year<sup>1</sup>. Specialized protein

transport systems—known as type VII secretion systems (T7SSs)—are central to the virulence of this pathogen, and are also crucial for nutrient and metabolite transport across the mycobacterial cell envelope<sup>2,3</sup>. Here we present the structure of an intact T7SS inner-membrane complex of *M. tuberculosis*. We show how the 2.32-MDa ESX-5 assembly, which contains 165 transmembrane helices, is restructured and stabilized as a trimer of dimers by the MycP<sub>5</sub> protease. A trimer of MycP<sub>5</sub> caps a central periplasmic dome-like chamber that is formed by three EccB<sub>5</sub> dimers, with the proteolytic sites of MycP<sub>5</sub> facing towards the cavity. This chamber suggests a central secretion and processing conduit. Complexes without MycP<sub>5</sub> show disruption of the EccB<sub>5</sub> periplasmic assembly and increased flexibility, which highlights the importance of MycP<sub>5</sub> for complex integrity. Beneath the EccB<sub>5</sub>–MycP<sub>5</sub> chamber, dimers of the EccC<sub>5</sub> ATPase assemble into three bundles of four transmembrane helices each, which together seal the potential central secretion channel. Individual cytoplasmic EccC<sub>5</sub> domains adopt two distinctive conformations that probably reflect different secretion states. Our work suggests a previously undescribed mechanism of protein transport and provides a structural scaffold to aid in the development of drugs against this major human pathogen.

[Download PDF](#)

## Main

*Mycobacterium tuberculosis* encodes five homologous, but functionally distinct, T7SSs that are designated ESX-1 to ESX-5. These systems translocate a number of effector proteins across the unique and impermeable diderm cell envelope<sup>2,3</sup>. Because of their importance for mycobacterial physiology and virulence, T7SSs are considered to be promising targets for the development of drugs for the treatment or prevention of tuberculosis<sup>4</sup>. Although T7SSs have previously been shown to form hexameric complexes<sup>5</sup>, high-resolution structural information exists only for part of the T7SS—a dimeric ESX-3 subcomplex from the nonpathogenic species *Mycobacterium smegmatis*<sup>6,7</sup>. Here we reconstituted the ESX-5 T7SS of *M. tuberculosis* H37Rv in *M. smegmatis* to obtain a structural view of the entire T7SS membrane complex from this human pathogen.

## Architecture and stoichiometry

The *M. tuberculosis* ESX-5 system showed robust expression in *M. smegmatis* and correct assembly of the membrane complex (Extended Data Fig. [1a, b](#)). Purification of the *M. tuberculosis* ESX-5 membrane complex (using a C-terminal Strep tag on EccC<sub>5</sub>

and mild solubilization conditions) resulted in copurification of the conserved MycP<sub>5</sub> protease (Extended Data Fig. 1c–e). To our knowledge, MycP (also known as mycosin) is absent in all previously reported T7SS structures<sup>5,6,7</sup>—although MycP is known to be essential for T7SS function and complex stability<sup>8</sup>. The addition of nucleotides and MgCl<sub>2</sub> improved sample homogeneity, as judged by a more-distinct high molecular weight complex on native-PAGE, size exclusion chromatography and subsequent negative-stain electron microscopy analysis (Extended Data Fig. 1f–j).

Cryo-electron microscopy (cryo-EM) analysis of the *M. tuberculosis* ESX-5 complex purified in the presence of ADP–AlF<sub>3</sub> showed clear hexameric particles (Extended Data Figs. 2b, 3). We performed an ab initio reconstruction without symmetry enforcement that yielded an average resolution of approximately 4 Å (Extended Data Fig. 4), which improved to an overall resolution of approximately 3.5 Å after further data processing; this allowed us to build around 78% of the stable complex de novo (Supplementary Tables 1, 2). The intact machinery comprises EccB<sub>5</sub>, EccC<sub>5</sub>, EccD<sub>5</sub>, EccE<sub>5</sub> and MycP<sub>5</sub> with a 6:6:12:6:3 stoichiometry (Fig. 1b, c, d, Supplementary Video 1) resulting in a 2.32-MDa complex that is anchored in the inner membrane through 165 transmembrane helices (TMHs) (Fig. 1e). The membrane assembly is best described as a trimer of dimers, in which each dimer comprises a single copy of MycP<sub>5</sub> and two protomers each of one copy of EccB<sub>5</sub>, EccC<sub>5</sub>, EccE<sub>5</sub> and two copies of EccD<sub>5</sub> (Fig. 1f). The overall fold and stoichiometry of a dimeric building block of *M. tuberculosis* ESX-5 is similar to that of a dimer of ESX-3 from *M. smegmatis*<sup>6,7</sup>, albeit with notable differences on the periplasmic side and the angle between protomers. In the intact ESX-5 complex, the angle between protomers at the membrane level differs by about 0.5°, from 59.7° between protomers of one dimer to 60.2° between protomers of adjacent dimers. However, at the cytosolic level these angles differ by more than 10° (from 65.3° to 54.7°) (Extended Data Fig. 5). By contrast, the individual ESX-3 dimer displayed an overall angle of 72° between protomers<sup>6,7</sup>.

**Fig. 1: Cryo-EM structure of the intact ESX-5 inner-membrane complex of *M. tuberculosis*.**

---

 **figure1**

---

**a**, Genetic organization of the *esx-5* locus of *M. tuberculosis* H37Rv, which was cloned and expressed in *M. smegmatis* MC<sup>2</sup>155. **b–e**, Cryo-EM density of the intact ESX-5 inner-membrane complex of *M. tuberculosis*, zoned and coloured for every individual component. Components are inner EccB<sub>5</sub> (dark green), outer EccB<sub>5</sub> (light green), EccC<sub>5</sub> (blue), inner EccD<sub>5</sub> (beige), outer EccD<sub>5</sub> (orange), EccE<sub>5</sub> (purple) and MycP<sub>5</sub> (red). The full complex is 28.5 nm in width and 20 nm in height, and has an absolute stoichiometry of 6:6:12:6:3 for EccB<sub>5</sub>:EccC<sub>5</sub>:EccD<sub>5</sub>:EccE<sub>5</sub>:MycP<sub>5</sub>. **b–e**, Side (**b**), top (**c**) and bottom (**d**) views and a top cross-section (**e**) of the complex at the membrane level, highlighting the arrangement of the 165-TMH region. Inset, top cross-section of an extracted dimeric unit. **f**, Single dimer viewed from the centre of the intact complex, highlighting the central EccC<sub>5</sub> TMH bundle and the position of MycP<sub>5</sub> with its active site directed towards the inside of the periplasmic cavity. **g**, Ribbon model of the *M. tuberculosis* ESX-5 assembly.

[Full size image](#)

## Structural rearrangements of periplasmic domains

The periplasmic assembly of the ESX-5 membrane complex of *M. tuberculosis* is formed by three EccB<sub>5</sub> dimers and three MycP<sub>5</sub> proteases. The EccB<sub>5</sub> dimers assemble in a triangle, which forms a central cavity (Fig. 2c). Within an EccB<sub>5</sub> dimer, two slightly different conformations (that is, inner and outer) can be distinguished between monomers, depending on their position. EccB<sub>5</sub> dimerization is mediated mainly through the R1 and R4 repeat domains and is further stabilized by the EccB<sub>5</sub> C termini, which wrap around their interacting EccB<sub>5</sub> partner to form intermolecular hydrophobic contacts with its R1 domain (Fig. 2d). The GIPGAP motif—which is a highly conserved region in EccB homologues—is central to these interactions (Fig. 2d). Compared to the EccB<sub>3</sub> dimer from the ESX-3 subassembly<sup>6,7</sup>, the three EccB<sub>5</sub> dimers are rotated by about 52° with respect to their corresponding EccC<sub>5</sub>–EccD<sub>5</sub>–EccE<sub>5</sub> membrane dimers, which indicates that large conformational rearrangements are required during maturation into the fully assembled hexamer (Extended Data Fig. 6d). To form the triangle-shaped assembly, the inner EccB<sub>5</sub> engages the outer EccB<sub>5</sub> of the adjacent dimer by packing its R3 domain against the  $\alpha$ -helices  $\alpha$ 5 and  $\alpha$ 8 of domains R2 and R3, respectively, which results in an asymmetric EccB<sub>5</sub>–tip arrangement (Fig. 2c). Consequently, domain R3 of the outer EccB<sub>5</sub> does not form any interactions at its tip extremity and thus displays higher flexibility, consistent with previous observations<sup>9</sup>.

**Fig. 2: MycP<sub>5</sub> drives EccB<sub>5</sub> hexamerization and stabilization of the membrane complex.**



**a**, Transparent assembly of intact *M. tuberculosis* ESX-5, with EccB<sub>5</sub> and MycP<sub>5</sub> coloured as in Fig. 1. **b**, Complete structure of monomeric *M. tuberculosis* EccB<sub>5</sub>, highlighting its overall fold and domains. **c**, Top and bottom view of the EccB<sub>5</sub>–MycP<sub>5</sub> periplasmic assembly with one unit (EccB<sub>5</sub> dimer and MycP<sub>5</sub> monomer) as ribbon model, highlighting the active site of MycP<sub>5</sub> in yellow. CD, central domain; TM, TMH. **d**, EccB<sub>5</sub> dimerization site, highlighting the C-terminus of outer EccB<sub>5</sub> that is wrapped around the R1 and R4 domains of the adjacent inner EccB<sub>5</sub> monomer, the conserved GIPGAP motif of EccB<sub>5</sub> (in yellow) and the interactions of the EccB<sub>5</sub> dimer with loop 2 and the linker connection of MycP<sub>5</sub>. **e**, Transparent map of *M. tuberculosis* ESX-5 without copurified MycP<sub>5</sub>, with EccB<sub>5</sub> highlighted in dark green. The high flexibility of EccB<sub>5</sub> and the overall heterogeneity of the membrane complex in the absence of MycP<sub>5</sub> is indicated by curved lines. **f**, EccB<sub>5</sub>–MycP<sub>5</sub> interaction surface, highlighting the three buried tryptophans. **g**, Angle variation range between protomers of the MycP<sub>5</sub>-bound (+) and two unbound (−) states (I and II). Intra, between two protomers within a dimer: 62.7°, 62.4° and 62.4° (MycP<sub>5</sub>-bound); 61.8°, 61.3° and 61.5° (unbound, I); 61.4°, 62.3° and 60.2° (unbound, II). Inter, between two protomers of adjacent dimers: 57.5°, 57.3° and 57.5° (MycP<sub>5</sub>-bound); 58.5°, 58.3° and 58.3° (unbound, I); and 57.9°, 58.3° and 59.6° (unbound, II).

[Full size image](#)

## Periplasmic MycP<sub>5</sub>–EccB<sub>5</sub> assembly

Our periplasmic *M. tuberculosis* ESX-5 map shows three MycP<sub>5</sub> proteases that form a dome-like structure, which cap the periplasmic central cavity (Fig. 2a,c, Supplementary Video 2). Interactions between EccB<sub>5</sub> and MycP<sub>5</sub> are mediated mainly by the MycP<sub>5</sub> protease domain and a composite interface that is generated by the R4 domain and loop 6 (residues Thr424 to Ser435) of the inner EccB<sub>5</sub> (Fig. 2f). The MycP<sub>5</sub>–EccB<sub>5</sub> interface covers a surface area of about 1,230 Å<sup>2</sup>, which leads to the burial of three conserved tryptophan residues (Trp437 and Trp469 of EccB<sub>5</sub>, and Trp523 of MycP<sub>5</sub>) (Fig. 2f). Additionally, loop 2 of MycP<sub>5</sub> binds to the C terminus of the outer EccB<sub>5</sub>, which explains why a deletion of this loop previously showed reduced ESX-5 secretion in *Mycobacterium marinum*<sup>10</sup> (Fig. 2d, Extended Data Fig. 6). MycP<sub>5</sub>–MycP<sub>5</sub> interactions are mediated mainly through loop 1 and the N-terminal extension (which run across the top of the MycP<sub>5</sub> protomers), and loop 3, which contacts the neighbouring protease domain from the side (Extended Data Fig. 7c,d). Loop 5 (residues Ala151 to Val271), which is cleaved during ESX-5

maturation<sup>11</sup>, folds along the interface of two protease domains towards the central pore formed by the MycP<sub>5</sub> trimer (Extended Data Fig. 7a). Although we could not build a complete model of loop 5 (owing to its high flexibility), this loop appears to cap the central periplasmic pore (Extended Data Fig. 7a). Notably, as loop 5 is not present in all mycosins and is dispensable for ESX-5 secretion<sup>10</sup>, a speculative role in gating remains to be identified. The active sites of the MycP<sub>5</sub> proteases face towards the central lumen of the cavity (Fig. 2c), which implies that potential substrates of this protease are translocated through—and processed within—this periplasmic chamber.

The dimer interface between the inner and outer EccB<sub>5</sub> is the largest in the periplasmic assembly, and covers a surface area of around 2,000 Å<sup>2</sup> and provides a solvation-free energy gain of  $\Delta G = -23$  kcal mol<sup>-1</sup> per dimer (Fig. 2d). By contrast, the interfaces formed between EccB<sub>5</sub> dimers each bury a surface area of about 600 Å<sup>2</sup>, with a cumulative energy gain of only  $\Delta G = -18$  kcal mol<sup>-1</sup> upon trimerization. This could provide an explanation as to why dimeric ESX subcomplexes are more stable than their fully assembled counterparts<sup>6,7</sup>. The intermolecular EccB<sub>5</sub>–MycP<sub>5</sub> interactions (which have a surface area of around 395 Å<sup>2</sup> and  $\Delta G = -0.1$  kcal mol<sup>-1</sup>) are even more modest, which provides a rationale for why interactions between MycP<sub>5</sub> and the membrane complex have so far remained unknown.

## MycP<sub>5</sub> stabilizes the entire membrane complex

To further investigate the effect of MycP<sub>5</sub> on the entire structure, we analysed MycP<sub>5</sub>-free *M. tuberculosis* ESX-5 complexes from the same preparation (Fig. 2e,g). These assemblies contained the same EccB<sub>5</sub>:EccC<sub>5</sub>:EccD<sub>5</sub>:EccE<sub>5</sub> stoichiometry as the fully assembled complexes (Extended Data Fig. 8). Following 3D reconstruction, we obtained two MycP<sub>5</sub>-free *M. tuberculosis* ESX-5 maps that displayed resolution estimates of about 4.5 and about 6.7 Å (Extended Data Fig. 4). The differences were most notable on the periplasmic side, on which the six EccB<sub>5</sub> copies showed high flexibility and did not form a stable triangular scaffold in the absence of MycP<sub>5</sub> (Fig. 2e). This shows that MycP<sub>5</sub> enables the trimerization of the EccB<sub>5</sub> dimers in the periplasm. This result is highly interesting, because mycosins are subtilisin-like proteases without any additional domains apart from a TMH, some small loops and an N-terminal extension that wraps around the protein<sup>12,13</sup>. A more structural role for mycosin has previously been predicted<sup>8</sup> but we now understand the essential role of mycosins in T7SSs.

In contrast to the periplasmic domain, the cytosolic and membrane regions of the MycP<sub>5</sub>-free maps were more similar to those of the MycP<sub>5</sub>-containing particles

(Extended Data Fig. 8). However, the MycP<sub>5</sub>-free particles displayed an increased heterogeneity that affected the entire complex, which resulted in a slight waving of the membrane region and an increased angle variation between individual protomers (Fig. 2g, Extended Data Fig. 8, Supplementary Video 4). The protease domain and TMH of MycP<sub>5</sub> synergistically reinforce the membrane complex. Their interactions with periplasmic inner EccB<sub>5</sub> and membrane-embedded outer EccD<sub>5</sub> (from separate protomers within a dimer) better anchor the periplasmic assembly to the membrane, while also stabilizing the dimeric unit (Extended Data Figs. 8f, 9). Additionally, by driving the formation of the periplasmic assembly, MycP<sub>5</sub> stabilizes the entire complex by promoting cross-dimer MycP<sub>5</sub>–MycP<sub>5</sub> and inner EccB<sub>5</sub>–outer EccB<sub>5</sub> interactions (Extended Data Fig. 9). Our MycP<sub>5</sub>-free *M. tuberculosis* ESX-5 maps are reminiscent of the hexameric, low-resolution negative-stain structure of ESX-5 from *Mycobacterium xenopi*, in which the periplasm was similarly disorganized in the absence of MycP<sub>5</sub>.

## EccC5 gates a potential secretion conduit

At the membrane level, six EccD<sub>5</sub>-dimer barrels (each of which contains 22 TMHs) together form a circular raft with an inner cavity (Extended Data Fig. 10). Within this raft, inner EccD<sub>5</sub> monomers are situated closer to the centre, whereas outer EccD<sub>5</sub> monomers face towards the periphery of the membrane complex. The EccD<sub>5</sub> membrane barrels are structurally highly similar to the homologous EccD<sub>3</sub> barrel in the ESX-3 subassembly<sup>6,7</sup>. The inner surface of each EccD<sub>5</sub> barrel is decorated with densities that are attributable to stably bound lipids or detergent molecules, which suggests that, in their native membrane environment, these barrels are filled with membrane lipids (Extended Data Fig. 10e,f).

The TMH of each copy of EccB<sub>5</sub> is anchored within the confinement of the EccD<sub>5</sub> raft through hydrophobic interfaces that are provided by TMH6 and TMH11 of inner EccD<sub>5</sub> and stably bound lipids (Fig. 3a, Extended Data Fig. 6f). Notably, no substantial intermolecular interactions can be found between adjacent EccD<sub>5</sub> barrels. Instead, coupling between two neighbouring EccD<sub>5</sub> barrels is achieved by the N-terminal loop and  $\alpha$ -helix of EccB<sub>5</sub> that run parallel to the cytoplasmic side of the inner membrane and engage in interactions with the TMHs of the neighbouring EccD<sub>5</sub> barrel in a clockwise manner (Fig. 3a, Extended Data Fig. 6g,h). Because the TMH of EccB<sub>5</sub> is slightly angled towards the centre of the complex, the architecture of the hexamer of EccB<sub>5</sub> TMHs is reminiscent of a basket, the inner diameter of which shrinks from around 60 Å to around 45 Å towards the periplasmic side (Fig. 3b).

**Fig. 3: A basket formed by the EccB<sub>5</sub> TMHs holds three four-TMH bundles of EccC<sub>5</sub>.**

 [figure3](#)

**a**, Angled view from the outside of the complex, showing the TMH and N terminus of an outer EccB<sub>5</sub> interacting with a pocket formed by TMH8, TMH10 and TMH11 of inner EccC<sub>5</sub> from the adjacent barrel. **b**, Side cross-section through the EccB<sub>5</sub> basket that contains the EccC<sub>5</sub> TMH bundles. Light blue densities depict the three copies of EccC<sub>5</sub> TMH2 that form the central pyramid. Two TMHs of EccC<sub>5</sub> were removed for clarity. Sizes indicate the inner diameters of the EccB<sub>5</sub> basket. **c**, Side cross-section through an EccB<sub>5</sub> basket, showing that the EccC<sub>5</sub> TMH bundle does not interact with outer EccB<sub>5</sub> from its own dimer, but instead forms lipid-mediated interactions with the outer EccB<sub>5</sub> TMH of the adjacent dimer. Lipids are shown in gold. **d**, Top view of the central EccB<sub>5</sub> basket and the EccC<sub>5</sub> TMH bundles. Dashed line marks the TMHs that belong to one dimeric unit. **e**, As in **d**, highlighting the lipid-rich environment. In the central area that surrounds the EccC<sub>5</sub> pyramid, lipids are not clearly distinguishable (which suggests fluidity in this area). **f**, **g**, Surface model displaying the hydrophobicity of an EccC<sub>5</sub> TMH bundle (**f**) and the EccB<sub>5</sub> basket (**g**). Hydrophilic amino acids are shown in turquoise, and hydrophobic residues are shown in sepia.

[Full size image](#)

EccC is the only component that is present in all T7SSs (including in related systems in Firmicutes<sup>2</sup>), and is therefore thought to be the central component in this nanomachinery. Each EccC protein has—in addition to two TMHs—four FtsK/SpoIIIE-like ATPase domains (also known as nucleotide-binding domains (NBDs)) that are known to be important for secretion<sup>6,7,14</sup>. We fully resolved the twelve EccC<sub>5</sub> TMHs in the intact *M. tuberculosis* ESX-5 complex; these form three four-TMH bundles, each of which belongs to the EccC<sub>5</sub> molecules of one dimer (Fig. 3b–d, Extended Data Fig. 11). These bundles are held together by hydrophobic interactions and effectively seal the central space of the membrane complex, which is enclosed by the EccB<sub>5</sub> basket (Fig. 3d, Supplementary Video 3). Two EccC<sub>5</sub> TMHs from each bundle contact the TMH of the inner EccB<sub>5</sub>, which leaves the outer EccB<sub>5</sub> TMH unbound by EccC<sub>5</sub> (Fig. 3c, d). At the very centre of the complex, one TMH of each bundle contributes to the formation of a pyramidal assembly that aligns with the periplasmic chamber (Extended Data Fig. 11d).

The chamber within the EccB<sub>5</sub> basket appears to be filled with lipids. However, the density for these lipids is more ambiguous than that of the lipids in and around the EccD<sub>5</sub> barrels, which suggests that the lipids within this chamber are more fluid (Fig. 3e). Notably, the local resolution gradually increases when moving from the centre to the EccB<sub>5</sub> basket, where the resolution is highest (Extended Data Fig. 11c). This indicates that the EccC<sub>5</sub> TMH bundles display more flexibility, as compared to the rigid EccB<sub>5</sub> basket. The entrance to the putative EccC<sub>5</sub> pore widens on the cytoplasmic side, where the EccC<sub>5</sub> stalk domains expand radially (Extended Data Figs. 5, 11a). Together, our data suggest that the six EccD<sub>5</sub> barrels provide a stable scaffold for assembly of a secretion pore that is confined by the EccB<sub>5</sub> TMHs and gated through three EccC<sub>5</sub> TMH bundles. Secretion through the inner membrane complex would require rearrangement of the EccC<sub>5</sub> TMHs. Such a proposed central pore would extend into the periplasmic chamber that is formed by EccB<sub>5</sub> and MycP<sub>5</sub>.

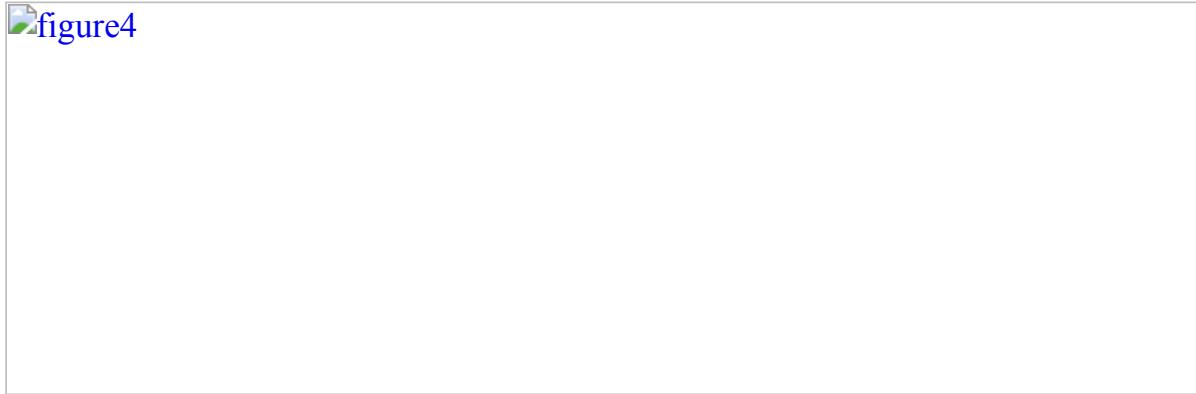
## Cytosolic EccC<sub>5</sub> adopts two conformations

At the cytoplasmic side of the complex, EccC<sub>5</sub> has a stalk helix that connects its second TMH to the first NBD (which is also known as the DUF domain). This NBD is bound to the cytosolic domains of inner and outer EccD<sub>5</sub>, which—in turn—are bound to EccE<sub>5</sub> at the periphery, together forming a ‘cytosolic bridge’ (Fig. 1f).

The distal C-terminal part of EccC<sub>5</sub>, which comprises a string of three NBDs (NBD1, NBD2 and NBD3), adopts two main conformations: we refer to these as extended and

contracted (Fig. 4a). In the extended state, the C-terminal three NBDs of EccC<sub>5</sub> bend parallel to the membrane, and align with the cytosolic domains of inner EccD<sub>5</sub> and EccE<sub>5</sub> of the same protomer, and extend beyond the diameter of the membrane assembly. Although of considerably lower resolution, this density can confidently accommodate a homology model that consists of the three EccC<sub>5</sub> NBD domains (Fig. 4a, Extended Data Fig. 12). Further classification of the extended state reveals EccC<sub>5</sub> to be more heterogeneous beyond NBD1, which suggests that NBD1 is more stably bound to components of its own protomer. Although we found only a relatively small number of particles in the contracted conformation, we solved the stable core of the membrane complex to sub-nanometre resolution (Extended Data Fig. 3). In the contracted state, the flexible arms of EccC<sub>5</sub> extend from the interface between the DUF domain of EccC<sub>5</sub> and the cytosolic domain of inner EccD<sub>5</sub> (Fig. 4a, Extended Data Fig. 12). We observed three separate disc-like structures that gradually constrict from the top to the bottom. This density shows a gap at the interface between NBD1 and NBD2. This would allow the previously postulated<sup>14,15</sup> binding of substrates to the linker 2 that connects NBD1 and NBD2, resulting in the displacement of this linker and the activation of NBD1. The highly dynamic cytoplasmic domains of the machinery may provide the basis for substrate selection, recognition or transport across the membrane.

**Fig. 4: EccC<sub>5</sub> adopts an extended and a contracted conformation.**



**a**, Side cross-section of density maps, showing the extended and contracted conformation of EccC<sub>5</sub>. The periplasmic and cytoplasmic chambers formed by EccB<sub>5</sub>–MycP<sub>5</sub> and by EccC<sub>5</sub> upon closing are highlighted. Homology models of the three C-terminal NBDs of EccC<sub>5</sub> are fitted in the cytosolic densities. Cytosolic bridge components are coloured as in Fig. 1. **b**, Model of the intact T7SS inner-membrane complex, highlighting the two conformations of EccC<sub>5</sub>.

[Full size image](#)

Our work provides a fully assembled structure of the ESX-5 inner-membrane complex of *M. tuberculosis*. As the membrane components of the five mycobacterial ESX systems show high sequence conservation, our results probably constitute a general structural blueprint for all of these T7SSs—including the virulence-related ESX-1 system (Supplementary Figs. 2, 3). Furthermore, our structure will serve as a platform for the identification of interactions that—if perturbed by small molecules—would aid in the treatment of tuberculosis.

## Methods

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.

### Molecular biology

*Escherichia coli* Dh5 $\alpha$  was grown at 37 °C and 200 rpm in LB medium supplemented with 30  $\mu$ g ml<sup>-1</sup> streptomycin. Cloning was performed in *E. coli* Dh5 $\alpha$  using iProof DNA polymerase from BioRad and restriction enzymes from New England Biolabs. A list of the primers used for amplification is available in Supplementary Table 3.

The plasmid expressing *M. tuberculosis* ESX-5 was built as follows: the backbone of the previously described pMV ESX-5<sub>mxen</sub> plasmid<sup>5</sup> was modified to encode the unique restriction sites DraI and PacI upstream and SpeI and NdeI downstream of the TwinStrep tag sequence. The *rv1782–rv1783* (*eccB*<sub>5</sub>–*eccC*<sub>5</sub>) region, including about 380 bp upstream of *eccB*<sub>5</sub>, of *M. tuberculosis* H37Rv was amplified while adding DraI and PacI restriction sites at the 5' and 3' ends, respectively (primers 1 and 2), and cloned into the modified plasmid upstream of the TwinStrep tag sequence, resulting in plasmid intermediate 1. The *M. tuberculosis* H37Rv region spanning *rv1791–rv1798* (*pe19–eccA*<sub>5</sub>) was amplified while adding SpeI and NdeI unique restriction sites (primers 3 and 4) and cloned downstream of the TwinStrep tag sequence into the intermediate 1, resulting in plasmid intermediate 2. Plasmid intermediate 2 was digested with SpeI and SnaBI, removing the region *rv1791–rv1794* (*pe19–espG*<sub>5</sub>), and the region encompassing *rv1785–rv1794* (*cyp143–espG*<sub>5</sub>) was amplified as two individual PCR products (primers 5 and 6 and primers 7 and 8). The restricted backbone and PCR products were InFusion (Takara Bio)-ligated, resulting in the final pMV-ESX-5<sub>mtb</sub> containing the entire *rv1782–rv1798* (*eccB*<sub>5</sub>–*eccA*<sub>5</sub>) locus.

### Isolation of mycobacterial cell envelopes

*Mycobacterium smegmatis* MC<sup>2</sup>155 expressing *M. tuberculosis* ESX-5 was grown at 37 °C and 90 rpm in LB medium supplemented with 0.05% Tween 80 and 30 µg ml<sup>-1</sup> streptomycin. Cultures were grown to an optical density (OD) at 600 nm of about 1.5, spun down for 15 min at 12,000g in a JLA-8.1000 rotor and subsequently washed in PBS. After culture collecting, all subsequent steps were performed at 4 °C. Washed cell pellets were resuspended in buffer A (50 mM Tris-HCl pH 8, 300 mM NaCl and 10% glycerol) at a concentration of about 50 OD ml<sup>-1</sup> and lysed by passing two times through a high-pressure homogenizer (Stansted) using a pressure of 0.83 kbar.

Unbroken cells were pelleted at 5,000g for 5 min and supernatants were transferred to ultracentrifugation tubes. Cell envelopes were separated from the soluble fraction by ultracentrifugation at 150,000g for 1.5 h. Following ultracentrifugation, supernatants were discarded, pellets were washed once with buffer A, resuspended in buffer A at a concentration of 750–1,000 OD ml<sup>-1</sup>, snap-frozen in liquid nitrogen and stored at –80 °C until further use. The protein concentration of the cell envelope fraction was measured by BCA assay (Pierce).

## Purification of the *M. tuberculosis* ESX-5 membrane complex

All steps were performed at 4 °C. The *M. tuberculosis* ESX-5 was purified as follows: cell envelope fractions were diluted to 3 mg ml<sup>-1</sup> in buffer B (50 mM Tris-HCl pH 8, 300 mM NaCl and 5% glycerol), supplemented with 0.25% DDM, 3 mM ADP-AlF<sub>3</sub> and 6 mM MgCl<sub>2</sub>. Following solubilization, the cell envelope mixture was spun down at 100,000g for 20 min, supernatants were collected and incubated with StrepTactin resin (IBA). Beads were subsequently washed with buffer B supplemented with 0.03% DDM, 1 mM ADP-AlF<sub>3</sub> and 2 mM MgCl<sub>2</sub>. Bound protein was eluted from the resin with buffer B supplemented with 0.03% DDM, 3 mM ADP-AlF<sub>3</sub>, 6 mM MgCl<sub>2</sub> and 10 mM desthiobiotin. The protein concentration of the eluate was measured by Bradford assay and amphipol A8-35 was added in an amphipol:protein ratio of 5:1. After a 1-h incubation, the amphipol-containing eluate was incubated overnight (around 12–16 h) with BioBeads in a BioBeads:detergent ratio of 20:1. Subsequently, BioBeads were removed using gravity flow chromatography columns and the sample was concentrated using Amicon Ultra 0.5-ml 100-kDa spin concentrators. The concentrated sample was further purified through size exclusion chromatography (SEC), using a Superose 6 Increase column running in buffer C (20 mM Tris-HCl pH 8, 200 mM NaCl) supplemented with 1 mM ADP-AlF<sub>3</sub> and 2 mM MgCl<sub>2</sub>. Size exclusion chromatography fractions were analysed by blue-native polyacrylamide gel electrophoresis (BN-PAGE) and negative-stain electron microscopy, after which the appropriate fractions were concentrated for cryo-EM using Amicon Ultra 0.5-ml 100-kDa spin concentrators. The initial Arctica sample was purified similarly, with the

addition of 5% glycerol in the SEC purification step and the omission of ADP–AlF<sub>3</sub> and MgCl<sub>2</sub> from the purification protocol.

## BN-PAGE

Samples consisting of either solubilized membranes or purified membrane complexes were mixed with 5% G-250 sample additive (Invitrogen), to a final concentration of about 0.2%, and run on 3–12% NativePage Bis-Tris Protein Gels (Invitrogen) according to manufacturer specifications. Gels were either stained with Coomassie R-250 or transferred to PVDF membranes and stained with appropriate antibodies (Supplementary Fig. 1). Antisera against EccB<sub>5</sub>, used in Extended Data Fig. 1b, was raised against the synthetic peptide CLPMMDMSPAELVVPK and has previously been described<sup>16</sup>. Polyclonal rabbit antisera against the peptide was raised in rabbits by Innovagen using Stimune (Prionix) as adjuvants. The antibody was used as a 1:5,000 dilution. Blots were visualized on a ChemoStar TouchMotionCor2 using ChemoStarTS.

## Negative-stain electron microscopy

Carbon-coated copper grids were glow-discharged for 30 s at 25 mA using a GloQube Plus Glow Discharge System (Electron Microscopy Sciences). Four microlitres of diluted sample was applied to the grids and incubated for 30 s. The sample was blotted off from the side and the grid was washed briefly with 4 µl of staining solution (2% uranyl acetate) and then stained with 4 µl of the staining solution for 30 s. The stain was blotted off from the side and grids were air-dried. Grids were imaged using a Thermo Fisher Scientific Talos L120C TEM equipped with a 4K Ceta CEMOS camera using TIA 4.1.5.

## Cryo-EM sample preparation

For the main datasets, purified sample was applied to Quantifoil R2/2, 200 mesh, copper grids floated with an additional approximately 1.1-nm layer of amorphous carbon. Four microlitres of sample was applied onto glow-discharged grids (30 s at 25 mA) and allowed to disperse for 60 s at 4 °C and 100% humidity. Grids were blotted for 4–6 s with a blot force of –5 and plunge-frozen in a liquid propane–ethane mixture, using a Thermo Fisher Scientific Vitrobot Mark V. For the initial Arctica dataset, all steps were similar, with the exception of the additional layer of amorphous carbon.

## Cryo-EM data acquisition

The initial cryo-EM dataset was collected on a 200-kV FEI Talos Arctica electron microscope equipped with a Falcon III direct electron detector running in counting mode and using Thermo Fisher Scientific EPU 1.11. A total of 853 movies were recorded with a nominal magnification of 150,000 $\times$ , corresponding to a pixel size of 0.96 at the specimen level. Movies were recorded with a total dose of 40.28 electrons per  $\text{Å}^2$ , fractionated in 38 frames over a 40-s exposure time and with a nominal defocus range of 1–2.5  $\mu\text{m}$ .

The two high-resolution datasets were recorded using Thermo Fisher Scientific EPU 2.4 software on a 300-kV Titan Krios TEM, equipped with a Gatan K3 direct electron detector running in counting mode and a Gatan Bioquantum energy filter (slit size 10 eV). We recorded 7,984 and 9,389 movies in counting mode in the two separate sessions with a nominal magnification of 81,000 $\times$ , corresponding to a pixel size of 1.1  $\text{Å}$  at the specimen level. Movies were recorded with a total dose of 59.5 electrons per  $\text{Å}^2$ , fractionated in 50 frames over a 3-s exposure time and with a nominal defocus range of 1–3  $\mu\text{m}$ .

## Cryo-EM data processing

Single-particle analysis was performed using Relion3.1<sup>17</sup>, unless stated otherwise. For the initial Arctica dataset, movies were motion-corrected using MotionCor2<sup>18</sup> and dose-weighted, and the contrast-transfer function (CTF) was estimated using CTFFIND4<sup>19</sup>. Automated particle-picking was performed using Cryolo<sup>20</sup> and the pretrained Janni model. Following particle extraction and several rounds of 2D classification to remove obvious artefacts, an initial de novo model was generated. The dataset was further cleaned using 3D classification and the best class was subsequently used for reference-based particle-picking. Following 2D and 3D classification (and 3D refinement in C1), the map displayed an apparent threefold symmetry and was further refined in C3. This final map displayed an estimated 13.5  $\text{Å}$  resolution.

In the first Krios dataset, movies were motion-corrected using MotionCor2, dose-weighted and the CTF was estimated using CTFFIND4. Automated particle-picking was performed using Cryolo with the pretrained Janni model and a low threshold. Particles were extracted and binned 4 $\times$  and several rounds of 2D classification were performed followed by 3D classification with the 30  $\text{Å}$ -filtered Arctica model as a template. The resulting particles were re-extracted without binning, CTF-corrected and polished and refined in C1, giving a map with an estimated overall resolution of 4.5  $\text{Å}$ . For the cytosolic region, particles were recentred on the cytosolic region, re-extracted, CTF-corrected, polished and 3D-refined. Following refinement, the density accounting for individual cytosolic dimers was subtracted, resulting in a particle stack that was three times larger. Cytosolic dimers were first refined with a mask encompassing both cytosolic bridges. Subsequently, these were focus-refined using a soft mask around

one of the cytosolic bridges. This map was refined using the default Relion value ‘--tau2fudge 2’ but also ‘--tau2fudge 4’, which increased the overall connectivity of the lower cytosolic area. The final map for the cytosolic bridge showed an estimated resolution of 3.3 Å and was sharpened using either Relion postprocessing or DeepEMhancer<sup>21</sup>. DeepEMhancer further helped to improve the observed anisotropy, overall map connectivity.

The second Krios dataset was processed similarly, with some exceptions. Following 3D classification of the binned data against the 4.5 Å Krios map filtered to 30 Å, the two maps with and without MycP<sub>5</sub> were processed separately. The MycP<sub>5</sub>-unbound map displayed increased heterogeneity and—following unbinned re-extraction and refinement—the particles were 3D-classified without alignment, resulting in two obvious classes of 4.5 Å and 6.7 Å resolution. Model free density modification in Phenix.Resolve\_Cryo\_EM<sup>22</sup> further improved the resolution to 4.3 Å and 5.8 Å, respectively. By contrast, a similar 3D classification for the MycP<sub>5</sub>-bound map did not further classify into structurally distinct classes. Following unbinned re-extraction and refinement, the MycP<sub>5</sub>-bound map showed an overall resolution of 4 Å, which was further improved to 3.8 Å after C3 refinement. Model free density modification in Phenix.Resolve\_Cryo\_EM further improved the resolution of the entire C1 map to 3.8 Å and of the C3 refined map to 3.56 Å. For the periplasmic map, the centre of mass for that region was determined using Chimera<sup>23</sup> and the particles were recentred, extracted, 3D-refined and polished to obtain the periplasmic map at an estimated 3.8 Å resolution in C1. Following 3D classification without alignment and further refinement in C3, the estimated resolution of the periplasmic map improved to 3.5 Å. To separate the two states of EccC<sub>5</sub>, particles were recentred on the lower cytosolic region, at the level of the DUF domain, polished and 3D-refined. This was followed by a masked 3D classification in which the mask contained NBD1 and NBD2 of EccC<sub>5</sub> in the extended conformation. The two main classes were further 3D-refined unmasked and subsequently masked, leading to a map of about 4.27 Å for the extended conformation and 7.6 Å for the contracted conformation.

## Model building and refinement

Model building started by generating homology models for MycP<sub>5</sub>, EccB<sub>5</sub>, EccC<sub>5</sub> and EccD<sub>5</sub> with Phyre2<sup>24</sup>. For MycP<sub>5</sub>, Protein Data Bank (PDB) entry 4J94<sup>12</sup> served as a structural template, and PDB entries 4KK7<sup>25</sup>, 4NH0<sup>14</sup> and 6SGW<sup>6</sup>, and 6SGZ<sup>6</sup> served as reference models for EccB<sub>5</sub>, EccC<sub>5</sub> and EccD<sub>5</sub>, respectively. To obtain atomic models of the periplasmic part (MycP<sub>5</sub>–EccB<sub>5</sub>) of the *M. tuberculosis* ESX-5 complex, homology models of MycP<sub>5</sub> and EccB<sub>5</sub> were rigid-body-fitted into a C1 symmetry, focus-refined periplasmic *M. tuberculosis* ESX-5 map (Electron Microscopy Data

Bank (EMDB) code EMD-12518) using the fit-in-map tool in ChimeraX (v.1.0)<sup>26</sup>. Model building, extension and interactive refinement was performed with ISOLDE (v.1.0.1)<sup>27</sup>, a molecular-dynamics-guided structure refinement tool within ChimeraX (v.1.0). The resulting coordinate file (PDB 7NPS) was further refined with Phenix.real\_space\_refine (v.1.18.2-3874)<sup>28</sup> using reference model restraints, strict rotamer matching and disabled grid search. Model validation was carried out using the MolProbity web server<sup>29</sup> and EMRinger<sup>30</sup> within the Phenix software package. Models for the membrane-embedded region (MycP<sub>5</sub>–EccB<sub>5</sub>–EccC<sub>5</sub>–EccD<sub>5</sub>) and cytoplasmic bridge (cytosolic domains of EccC<sub>5</sub>–EccD<sub>5</sub>) (PDB 7NPT) were built in the same way, using a reconstruction of the full *M. tuberculosis* ESX-5 complex (EMDB EMD-12517) and a focus-refined map of the cytoplasmic domains (EMDB EMD-12520) sharpened with DeepEMhancer, respectively. Finally, a composite model was assembled by fusing the periplasmic assembly and six copies of the cytosolic bridge to the membrane-embedded region model. This composite model was then refined against the full *M. tuberculosis* ESX-5 complex map with C1 symmetry (PDB 7NP7 and EMDB EMD-12514) and C3 symmetry (PDB 7NPR and EMDB EMD-12517).

Modelling into MycP<sub>5</sub>-free maps was performed with ISOLDE using the composite ESX-5 model, in which MycP<sub>5</sub> and the periplasmic domain of EccB<sub>5</sub> (residues 84–507) had been deleted. Adaptive distance restraints as well as torsion restraints were applied to all atoms to restrain short-range conformational changes but allow for long-range conformational movements. ISOLDE simulations for dynamic fitting of the coordinate file into EMD-12521 and EMD-12522 were performed (about 10 min, 10 K) after which the models showed satisfying fits to the new maps without further manual intervention. MycP<sub>5</sub>-free models were further refined against the maps using Phenix.real\_space\_refine (v.1.18.2-3874) as stated.

Visualization of atomic coordinates and map volumes was performed with ChimeraX (v.1.1) and PyMOL v.2.4.0<sup>31</sup>. Buried surface areas between subunits were calculated by PISA<sup>32</sup>.

## Reporting Summary

Further information on research design is available in the [Nature Research Reporting Summary](#) linked to this article.

## Data availability

Cryo-EM maps have been deposited in the EMDB under accession codes [EMD-12514](#) (full complex in C1), [EMD-12517](#) (full complex in C3), [EMD-12518](#) (periplasmic map in C1), [EMD-12519](#) (periplasmic map in C3), [EMD-12520](#) (cytosolic bridge), [EMD-12521](#) (MycP<sub>5</sub>-free map 1), [EMD-12522](#) (MycP<sub>5</sub>-free map 2), [EMD-12523](#) (EccC<sub>5</sub>, extended state) and [EMD-12525](#) (EccC<sub>5</sub>, contracted state). The composite model settled in the C1 and C3 full maps, periplasm in C1, cytosolic bridge, MycP<sub>5</sub>-free map 1 and MycP<sub>5</sub>-free map 2 have been deposited in the PDB under accession codes [7NP7](#), [7NPR](#), [7NPS](#), [7NPT](#), [7NPU](#) and [7NPV](#), respectively. All other data are available from the corresponding author upon reasonable request.

## References

1. 1.

WHO. *Global Tuberculosis Report* (WHO, 2020).

2. 2.

Bunduc, C. M., Bitter, W. & Houben, E. N. G. Structure and function of the mycobacterial type VII secretion systems. *Annu. Rev. Microbiol.* **74**, 315–335 (2020).

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

3. 3.

Gröschel, M. I., Sayes, F., Simeone, R., Majlessi, L. & Brosch, R. ESX secretion systems: mycobacterial evolution to counter host immunity. *Nat. Rev. Microbiol.* **14**, 677–691 (2016).

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

4. 4.

Rybníkář, J. et al. Anticytolytic screen identifies inhibitors of mycobacterial virulence protein secretion. *Cell Host Microbe* **16**, 538–548 (2014).

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

5. 5.

Beckham, K. S. et al. Structure of the mycobacterial ESX-5 type VII secretion system membrane complex by single-particle analysis. *Nat. Microbiol.* **2**, 17047

(2017).

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

6. 6.

Famelis, N. et al. Architecture of the mycobacterial type VII secretion system. *Nature* **576**, 321–325 (2019).

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

7. 7.

Powelet, N. et al. The structure of the endogenous ESX-3 secretion system. *eLife* **8**, e52983 (2019).

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

8. 8.

van Winden, V. J. C. et al. Mycosins are required for the stabilization of the ESX-1 and ESX-5 type VII secretion membrane complexes. *MBio* **7**, e01471-16 (2016).

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

9. 9.

Xie, X. Q. et al. Crystallographic observation of the movement of the membrane-distal domain of the T7SS core component EccB<sub>1</sub> from *Mycobacterium tuberculosis*. *Acta Crystallogr. F* **72**, 139–144 (2016).

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

10. 10.

van Winden, V. J. C., Damen, M. P. M., Ummels, R., Bitter, W. & Houben, E. N. G. Protease domain and transmembrane domain of the type VII secretion mycosin protease determine system-specific functioning in mycobacteria. *J. Biol. Chem.* **294**, 4806–4814 (2019).

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

11. 11.

van Winden, V. J. C., Bunduc, C. M., Ummels, R., Bitter, W. & Houben, E. N. G. A chimeric EccB–MycP fusion protein is functional and a stable component of the ESX-5 type VII secretion system membrane complex. *J. Mol. Biol.* **432**, 1265–1278 (2020).

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

12. 12.

Solomonson, M. et al. Structure of the mycosin-1 protease from the mycobacterial ESX-1 protein type VII secretion system. *J. Biol. Chem.* **288**, 17782–17790 (2013).

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

13. 13.

Wagner, J. M. et al. Understanding specificity of the mycosin proteases in ESX/type VII secretion by structural and functional analysis. *J. Struct. Biol.* **184**, 115–128 (2013).

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

14. 14.

Rosenberg, O. S. et al. Substrates control multimerization and activation of the multi-domain ATPase motor of type VII secretion. *Cell* **161**, 501–512 (2015).

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

15. 15.

Bunduc, C. M., Ummels, R., Bitter, W. & Houben, E. N. G. Species-specific secretion of ESX-5 type VII substrates is determined by the linker 2 of EccC<sub>5</sub>. *Mol. Microbiol.* **114**, 66–76 (2020).

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

16. 16.

Houben, E. N. et al. Composition of the type VII secretion system membrane complex. *Mol. Microbiol.* **86**, 472–484 (2012).

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

17. 17.

Zivanov, J., Nakane, T. & Scheres, S. H. W. Estimation of high-order aberrations and anisotropic magnification from cryo-EM data sets in *RELION-3.1*. *IUCrJ* **7**, 253–267 (2020).

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

18. 18.

Zheng, S. Q. et al. MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat. Methods* **14**, 331–332 (2017).

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

19. 19.

Rohou, A. & Grigorieff, N. CTFFIND4: fast and accurate defocus estimation from electron micrographs. *J. Struct. Biol.* **192**, 216–221 (2015).

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

20. 20.

Wagner, T. & Raunser, S. The evolution of SPHIRE-crYOLO particle picking and its application in automated cryo-EM processing workflows. *Commun. Biol.* **3**, 61 (2020).

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

21. 21.

Sanchez-Garcia, R. et al. DeepEMhancer: a deep learning solution for cryo-EM volume post-processing. Preprint at <https://doi.org/10.1101/2020.06.12.148296> (2020).

22. 22.

Terwilliger, T. C., Ludtke, S. J., Read, R. J., Adams, P. D. & Afonine, P. V. Improvement of cryo-EM maps by density modification. *Nat. Methods* **17**, 923–927 (2020).

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

23. 23.

Pettersen, E. F. et al. UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).

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

24. 24.

Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N. & Sternberg, M. J. The Phyre2 web portal for protein modeling, prediction and analysis. *Nat. Protocols* **10**, 845–858 (2015).

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

25. 25.

Wagner, J. M. et al. Structures of EccB1 and EccD1 from the core complex of the mycobacterial ESX-1 type VII secretion system. *BMC Struct. Biol.* **16**, 5 (2016).

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

26. 26.

Pettersen, E. F. et al. UCSF ChimeraX: structure visualization for researchers, educators, and developers. *Protein Sci.* **30**, 70–82 (2021).

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

27. 27.

Croll, T. I. ISOLDE: a physically realistic environment for model building into low-resolution electron-density maps. *Acta Crystallogr. D* **74**, 519–530 (2018).

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

28. 28.

Liebschner, D. et al. Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix. *Acta Crystallogr. D* **75**, 861–877 (2019).

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

29. 29.

Williams, C. J. et al. MolProbity: more and better reference data for improved all-atom structure validation. *Protein Sci.* **27**, 293–315 (2018).

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

30. 30.

Barad, B. A. et al. EMRinger: side chain-directed model and map validation for 3D cryo-electron microscopy. *Nat. Methods* **12**, 943–946 (2015).

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

31. 31.

The PyMOL Molecular Graphics System v.1.8 (Schrödinger, 2015).

32. 32.

Krissinel, E. & Henrick, K. Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol.* **372**, 774–797 (2007).

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

[Download references](#)

## Acknowledgements

We thank all members of the laboratories of T.C.M., E.N.G.H. and W.B. for their support of this project; W. Lugmayr for scientific IT support; L. Ciccarelli for initial cryo-EM analysis of the *M. tuberculosis* ESX-5 complex; and T. Croll for his support with ISOLDE. High-performance computing was possible through access to the HPC at DESY/Hamburg. Cryo-EM data collection was performed at the Cryo-EM Facility at CSSB. This project was supported by funds available to T.C.M. through the Behörde für Wissenschaft, Forschung und Gleichstellung of the city of Hamburg at the Institute of Structural and Systems Biology at the University Medical Center Hamburg–Eppendorf (UKE). The laboratory of T.C.M. is supported by DESY (German Electron Synchrotron Center). The cryo-EM facility is supported by the University of Hamburg, the University Medical Center Hamburg–Eppendorf and DFG grant numbers INST152/772-1, 152/774-1, 152/775-1, 152/776-1 and 152/777-1 FUGG. This work received funding by a VIDI grant (864.12.006; to C.M.B. and E.N.G.H.) from the Netherlands Organization of Scientific Research. This project has received funding from the European Union’s Horizon 2020 research and innovation

programme under the Marie Skłodowska-Curie grant agreement no. 101030373 (to C.M.B.).

## Funding

Open access funding provided by Deutsches Elektronen-Synchrotron (DESY).

## Author information

### Affiliations

1. Centre for Structural Systems Biology, Hamburg, Germany

Catalin M. Bunduc, Dirk Fahrenkamp, Jiri Wald & Thomas C. Marlovits

2. Institute of Structural and Systems Biology, University Medical Centre Hamburg–Eppendorf, Hamburg, Germany

Catalin M. Bunduc, Dirk Fahrenkamp, Jiri Wald & Thomas C. Marlovits

3. Deutsches Elektron Synchrotron DESY, Hamburg, Germany

Catalin M. Bunduc, Dirk Fahrenkamp, Jiri Wald & Thomas C. Marlovits

4. Molecular Microbiology Section, Amsterdam Institute of Molecular and Life Sciences, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands

Catalin M. Bunduc, Wilbert Bitter & Edith N. G. Houben

5. Department of Medical Microbiology and Infection Control, Amsterdam Infection and Immunity Institute, Amsterdam UMC, Amsterdam, The Netherlands

Roy Ummels & Wilbert Bitter

### Authors

1. Catalin M. Bunduc

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

2. Dirk Fahrenkamp

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

3. Jiri Wald

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

4. Roy Ummels

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

5. Wilbert Bitter

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

6. Edith N. G. Houben

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

7. Thomas C. Marlovits

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

## Contributions

C.M.B., D.F., J.W., E.N.G.H., W.B. and T.C.M. designed experiments. R.U. and C.M.B. generated constructs. C.M.B. purified complexes and performed biochemical assays. J.W. vitrified samples and collected cryo-EM images. C.M.B. and J.W. collected negative-stain images. D.F. built the atomic model. C.M.B., D.F., E.N.G.H., W.B. and T.C.M. interpreted data. C.M.B. and T.C.M. processed cryo-EM data. C.M.B., D.F., E.N.G.H., W.B. and T.C.M. wrote and revised the paper. All authors read, corrected and approved the manuscript. E.N.G.H., W.B. and T.C.M. supervised the project.

## Corresponding author

Correspondence to [Thomas C. Marlovits](#).

## Ethics declarations

### Competing interests

The authors declare no competing interests.

### Additional information

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

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

### Extended data figures and tables

#### [Extended Data Fig. 1 Purification of the \*M. tuberculosis\* ESX-5 membrane complex.](#)

**a**, Genetic organization of the *esx-5* locus of *M. tuberculosis* H37Rv, which has been cloned and expressed in *M. smegmatis* MC<sup>2</sup>155. **b**, BN-PAGE and western blot analysis using an anti-EccB<sub>5</sub> antibody of DDM-solubilized membranes from *M. smegmatis* MC<sup>2</sup>155 expressing *M. xenopi* or *M. tuberculosis* ESX-5 (ESX-5<sub>mxen</sub> or ESX-5<sub>mtb</sub>, respectively). Experiment was reproduced three times. **c**, **d**, Coomassie-stained SDS-PAGE (**c**) and BN-PAGE (**d**) of Strep- and SEC-purified ESX-5<sub>mtb</sub> membrane complexes. **e**, Negative-stain electron microscopy analysis of ESX-5<sub>mtb</sub> membrane complexes shown in **c** and **d**. Experiments in **c–e** were replicated three times. **f**, BN-PAGE and Coomassie staining of Strep-purified ESX-5<sub>mtb</sub> complexes without nucleotides (–) or in the presence of nucleotides ATP, ADP or the transition-state analogue ADP-AlF<sub>3</sub>. Upon purification, in the presence of either nucleotide, the higher-molecular-weight species of the membrane complex becomes more prominent. **g**, SDS-PAGE and Coomassie staining of the same samples as in **f**, showing a similar SDS-PAGE protein pattern between the four conditions. Experiment in **f, g** was performed three times. **h**, **i**, Coomassie-stained SDS-PAGE (**h**) and BN-PAGE (**i**) of Strep- and SEC-purified ESX-5<sub>mtb</sub> membrane complexes in the presence of ADP-AlF<sub>3</sub>. **j**, Negative-stain electron microscopy of the same sample as in **h, i**, showing improved sample homogeneity as compared to purifications in the absence of nucleotides, as shown in **e**. Experiment shown in **h–j** was performed twice.

## [Extended Data Fig. 2 Cryo-EM data collection and single-particle reconstruction procedure.](#)

**a, b**, This figure relates to the initial Talos-Arctica-collected dataset (**a**) and the first higher-resolution Titan-Krios-collected dataset (**b**).

## [Extended Data Fig. 3 Cryo-EM data collection and single-particle reconstruction procedure.](#)

This figure relates to the second high-resolution Titan Krios collected dataset.

## [Extended Data Fig. 4 Single-particle reconstructions of the ESX-5<sub>mtb</sub> membrane complex.](#)

**a–c**, Angular distribution plots, local-resolution estimations and Fourier shell correlation (FSC) plots of the C1 reconstruction of the entire MycP<sub>5</sub>-bound ESX-5<sub>mtb</sub> membrane complex (**a**), and C1 reconstructions of the two heterogeneous MycP<sub>5</sub>-unbound ESX-5<sub>mtb</sub> membrane complexes (**b, c**). **d, e**, Local-resolution estimation and FSC plot for the C1-refined periplasmic map (**d**) and the map of the cytosolic bridge (**e**). **f**, Examples of cryo-EM densities and corresponding models.

## [Extended Data Fig. 5 Top cross-sections through the intact ESX-5<sub>mtb</sub> membrane complex.](#)

**a**, MycP<sub>5</sub> trimer top view, highlighting the pore formed at the periplasmic side. **b**, Section through the periplasmic assembly at the EccB<sub>5</sub>–MycP<sub>5</sub> interface, showing the position of the protease domain sitting on top of inner EccB<sub>5</sub>. **c**, Section through the periplasmic assembly at the EccB<sub>5</sub> dimer interface level, highlighting the MycP<sub>5</sub> linker connection to the TMH. **d**, Top view of the six membrane protomers with the closed EccC<sub>5</sub> TMH pyramid at the centre. **e**, Top cross-section through the six membrane protomers, highlighting 153 of the 165 TMHs. At the central area towards the cytosol, the three-EccC<sub>5</sub> TMH pyramid opens up in a manner similar to an iris. MycP<sub>5</sub>, the protease domain of which interacts with the protomer containing inner EccB<sub>5</sub>, interacts with the outer EccD<sub>5</sub> barrel of the adjacent protomer at the membrane level. At the membrane level, the angle between protomers within a dimer and between adjacent protomers of different dimers differs by only 0.5°. **f**, Top section displaying the region below the inner leaflet of the inner membrane, highlighting a further opening of the EccC<sub>5</sub> gated pore and the lower part of the EccB<sub>5</sub> basket,

formed by EccB<sub>5</sub> N termini. **g**, At the cytosolic level, the angle between protomers differs to that at the membrane level. As such, the angle between protomers within a dimer grows to 65.3°, while the angle between adjacent protomers of different dimers decreases to 54.7°. The change in angles between the membrane and cytosolic regions of protomers is caused by MycP<sub>5</sub> binding, which induces a slight tilting to the protomers that it binds via inner EccD<sub>5</sub>. **h**, Section through the lower region of the cytosolic bridge, containing the DUF domain of EccC<sub>5</sub> and the cytosolic domain of inner EccD<sub>5</sub>. **i**, Same view as in **h**, but overlaid with the EccC<sub>5</sub> extended state, highlighting the radial extension of the EccC<sub>5</sub> NBD1, NBD2 and NBD3 almost parallel to the inner membrane. **j**, Same view as in **h**, but then overlaid with the EccC<sub>5</sub> contracted state.

### Extended Data Fig. 6 Hexameric EccB<sub>5</sub> adopts a triangular conformation in the periplasm.

**a, b**, Side (**a**) and top (**b**) view of an intact ESX-5<sub>mtb</sub> assembly in which inner EccB<sub>5</sub> and outer EccB<sub>5</sub> are coloured as in Fig. 1 and the rest of the components are transparent. **c**, A V-shaped EccB<sub>3</sub> dimer (PDB 6SGY) was fitted into the *M. smegmatis* ESX-3 dimer cryo-EM density (EMDB EMD-20820) together with the corresponding dimeric ESX-3 model for the membrane and cytosolic domains (PDB 6UMM) using the Chimera fit in map tool. This composite dimer model was subsequently trimerized, on the basis of our full ESX-5 map reconstructions. The clashing of EccB<sub>3</sub> periplasmic domains between the dimers, towards the central area, in this hybrid model are highlighted in red. **d**, Upon MycP<sub>5</sub> binding to the assembly, the periplasmic EccB<sub>5</sub> dimer is rotated by 52°, avoiding the clashes observed in **c**. Angles were measured by aligning the hybrid model and the ESX-5<sub>mtb</sub> model at the membrane level. Subsequently, centres of mass were defined for the combined R1 domains of each EccB<sub>3</sub> and EccB<sub>5</sub> dimer (at the base of the dimer) and for every R2 and R3 EccB monomer (toward the tips of the EccB dimer). Planes defined by these three points were generated for both EccB<sub>3</sub> and EccB<sub>5</sub> dimers and angles were measured between these two planes. EccB<sub>3</sub> dimer is shown as a ribbon model and EccB<sub>5</sub> model is shown as zoned density. **e**, Compared to the V-shaped EccB<sub>3</sub> dimer (ribbon model), the angle between the two EccB<sub>5</sub> monomers (zoned density) grows by 48° upon MycP<sub>5</sub> binding. EccB dimer angles were calculated by measuring the angle between the centres of mass of the R2 and R3 domains of each EccB protomer in relation to the centre of mass of both R1 domains. **f**, Side views of the TMH region of inner EccD<sub>5</sub>, depicted as a ribbon model, and the TMH and N terminus of an interacting outer EccB<sub>5</sub>, depicted as zoned density. An array of lipids found in the EccD<sub>5</sub> barrel (but also

surrounding this inner EccD<sub>5</sub>–EccB<sub>5</sub> interaction site) are depicted in gold. **g**, Bottom view of the lower cytosolic area of the EccB<sub>5</sub> basket, formed by EccB<sub>5</sub> N termini (residues 10–48) and depicted with the interacting pocket formed by TMH10, TMH11 and TMH8 (not shown for clarity) of inner EccD<sub>5</sub> of the adjacent protomer. The EccB<sub>5</sub> N terminus is also buttressed in this position by a short helix (residues 119–130) of outer EccD<sub>5</sub>, which connects outer EccD<sub>5</sub> TMHs with its cytosolic domain, and also by part of the inner EccD<sub>5</sub> loop (residues 307–315) that subsequently folds along the stalk and DUF domain of EccC<sub>5</sub>. **h**, Same map as in **g**, but viewed from the top. **k**, Superposition of inner EccB<sub>5</sub> and outer EccB<sub>5</sub>, highlighting conformational differences between the two, which are the result of the interaction with MycP<sub>5</sub>.

### **Extended Data Fig. 7 MycP<sub>5</sub> caps a periplasmic cavity with its active site directed towards the lumen.**

**a**, Side and top view of an intact ESX-5<sub>mtb</sub> assembly with MycP<sub>5</sub> coloured as in Fig. 1 and the rest of the components transparent. Insets, side and top views of the periplasmic assembly with EccB<sub>5</sub> in white, the MycP<sub>5</sub> density shown in transparent red and loop 5 of MycP<sub>5</sub> depicted in solid red at a higher threshold, to highlight it capping the periplasmic pore. Loop 5 folds along the protease domain, towards the pore formed by the MycP<sub>5</sub> trimer. At higher thresholds, loop 5 caps this pore. **b**, Top and side view of a dimer of EccD<sub>5</sub> barrels, of which one barrel (left) binds via inner EccD<sub>5</sub> to the MycP<sub>5</sub> TMH. **c**, Top or bottom view of MycP<sub>5</sub> trimers depicted in grey with the loops that are involved in MycP<sub>5</sub>–MycP<sub>5</sub> interactions depicted in different colours. **d**, Side and bottom views showing the MycP<sub>5</sub>–MycP<sub>5</sub> interactions mediated by the same domains depicted in the same colours as in **e**.

### **Extended Data Fig. 8 MycP<sub>5</sub> drives hexamerization of periplasmic EccB<sub>5</sub> and complex stability.**

**a**, Cryo-EM density map of a MycP<sub>5</sub>-free ESX-5<sub>mtb</sub> membrane complex, zoned and coloured as in Fig. 1. In the absence of MycP<sub>5</sub>, the periplasmic domains of EccB<sub>5</sub> display high flexibility. The rest of the membrane complex displays increased heterogeneity when compared to the MycP<sub>5</sub>-bound map. **b**, Map of difference created by subtracting the MycP<sub>5</sub>-free map from the MycP<sub>5</sub>-bound map. **c**, Overlay of **a** and **b**. **d**, MycP<sub>5</sub>-bound map in red and the two MycP<sub>5</sub>-free maps in blue and green. **e**, A model of the MycP<sub>5</sub>-bound map, in which MycP<sub>5</sub> and residues 84–504 of EccB<sub>5</sub> were removed, was fitted into the models of the two MycP<sub>5</sub>-free maps, as described in

Methods. Models were aligned at one EccD<sub>5</sub> barrel (dark dotted circle), revealing substantial variations and shifts between the three maps. Top inset shows that there is consistent variation between all three maps at the membrane level (EccD<sub>5</sub> barrel). Middle inset shows variations between maps at cytosolic level (EccB<sub>5</sub> N-terminal helix, residues 20–38). Bottom inset highlights inner EccD<sub>5</sub> from the EccD<sub>5</sub> barrel that was used for the alignment, showing that overall protomer structure does not change in the absence of MycP<sub>5</sub>. **f**, Dimers from every individual map, colour-coded the same as in **d** (different shades), were extracted and aligned to each other on one EccD<sub>5</sub> barrel (left) as in **e**. Insets from these alignments, derived from both protomers, show that all three dimers of the MycP<sub>5</sub>-bound map show little to no variation, whereas the two MycP<sub>5</sub>-free maps show a higher degree of heterogeneity between dimers.

### Extended Data Fig. 9 MycP<sub>5</sub> creates more interaction points between protomers and dimers.

**a**, Transparent surface model of a MycP<sub>5</sub>-free map with one EccB<sub>5</sub> dimer at the periplasmic side, highlighting the interfaces of the protomers from a dimer. In the absence of MycP<sub>5</sub>, the two protomers from within one dimer exhibit two interactions: an EccB<sub>5</sub>–EccB<sub>5</sub> interaction between their periplasmic domains, and a cytosolic one between EccB<sub>5</sub>–inner EccD<sub>5</sub>. The dimer further contacts the two immediate protomers of adjacent dimers through the mentioned EccB<sub>5</sub>–EccD<sub>5</sub> cytosolic interaction. **b**, Transparent surface model of the MycP<sub>5</sub>-bound map, highlighting the interface of protomers from a dimer. On top of the mentioned contacts, in the presence of MycP<sub>5</sub>, protomers from a dimer interact with each other at the periplasmic side through inner EccB<sub>5</sub>–MycP<sub>5</sub>–outer EccB<sub>5</sub>, while MycP<sub>5</sub> further anchors the periplasmic assembly to the stable EccD<sub>5</sub> raft through EccB<sub>5</sub>–MycP<sub>5</sub>–inner EccD<sub>5</sub>. MycP<sub>5</sub> also guides dimer–dimer interactions. By stabilizing the three EccB<sub>5</sub> dimers in the triangle assembly, MycP<sub>5</sub> promotes inner EccB<sub>5</sub>–outer EccB<sub>5</sub> interactions between opposing protomers from adjacent dimers. Additionally, MycP<sub>5</sub> promotes dimer–dimer contacts through MycP<sub>5</sub>–MycP<sub>5</sub> interactions in the periplasm. Colour-coded legend applies to **a**, **b**, **c**, **d**. Inside (**c**) and outside (**d**) view of a dimer containing EccB<sub>5</sub>, MycP<sub>5</sub> and the TMHs of EccD<sub>5</sub>. For purposes of clarity, one EccD<sub>5</sub>–EccB<sub>5</sub> protomer is coloured in blue, and the second protomer is in green and the MycP<sub>5</sub> in red. Interactions between protomers of a dimer are highlighted and colour-coded as in **a**, **b**. **e**, Top view of a surface model missing the periplasmic domains of EccB<sub>5</sub> and MycP<sub>5</sub>. The planes of protomers in which MycP<sub>5</sub> binds inner EccD<sub>5</sub> are tilted by about 5° compared to the MycP<sub>5</sub>-unbound ones.

## Extended Data Fig. 10 Six lipid-filled EccD<sub>5</sub> barrels form a central raft.

**a**, Side and top view of an intact ESX-5<sub>mtb</sub> assembly, in which inner EccD<sub>5</sub> and outer EccD<sub>5</sub> are coloured as in Fig. 1 and the rest of the components are transparent. **b**, Top view of the membrane region of the ESX-5<sub>mtb</sub> model overlaid with observed lipids, coloured in bright yellow. **c**, Same view as **b**, but showing only the lipids. **d**, Same image as **c**, but rotated 90° to show a side view, highlighting a bilayer-like structure. **e**, Top view of an EccD<sub>5</sub> barrel with observed lipids bound to inner EccD<sub>5</sub>. **f**, Side view of an inner EccD<sub>5</sub> monomer displayed as zoned density and overlaid with observed lipids.

## Extended Data Fig. 11 Three four-TMH-bundles of EccC<sub>5</sub> gate a central pore.

**a**, Side and top view of an intact ESX-5<sub>mtb</sub> assembly, in which EccC<sub>5</sub> is coloured in alternating light and dark blue and the rest of the components are transparent. **b**, Extracted dimeric EccC<sub>5</sub> and the TMHs and N termini of EccB<sub>5</sub> from the same dimer. The 90° inset rotation shows that the TMH of outer EccB<sub>5</sub> is not contacted by the TMHs of EccC<sub>5</sub>. **c**, Top membrane cross-section through a local-resolution map of a C1 full-complex reconstruction, displaying decreased resolution of the central space occupied by the TMHs of EccC<sub>5</sub>, compared to the surrounding EccB<sub>5</sub> basket and TMHs of inner EccD<sub>5</sub>. **d**, Top view of the full membrane complex with the EccC<sub>5</sub> TMH pyramid in light blue and the periplasmic EccB<sub>5</sub>–MycP<sub>5</sub> in the same colours as in Fig. 1. The EccC<sub>5</sub> TMH pyramid aligns with the periplasmic cavity and the MycP<sub>5</sub>-formed pore. MycP<sub>5</sub> top part is partially sectioned, for clarity. Inset showing a 90° rotation side cross-section of the same map. **e**, Ribbon model highlighting the structural features of the cytosolic bridge.

## Extended Data Fig. 12 EccC<sub>5</sub> adopts two separate conformations.

**a**, Extended conformation in which an EccC<sub>5</sub> NBD1–NBD2–NBD3 model is fitted to highlight the overall position of these domains with respect to the rest of the membrane complex. **b**, As in **a**, but for the contracted conformation.

## **Supplementary information**

### Supplementary Information

This file contains Supplementary Tables 1-3 and Supplementary Figures 1-3.

## Reporting Summary.

## Peer Review File

### Video 1

Overall structure of the ESX-5mtb membrane complex.

### Video 2

The periplasmic dome of the ESX-5mtb membrane complex.

### Video 3

The central, EccC5 gated, pore of the ESX-5mtb membrane complex.

### Video 4

Flexibility of the MycP5 free maps and stabilization upon MycP5 binding.

## 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

Bunduc, C.M., Fahrenkamp, D., Wald, J. *et al.* Structure and dynamics of a mycobacterial type VII secretion system. *Nature* **593**, 445–448 (2021).  
<https://doi.org/10.1038/s41586-021-03517-z>

[Download citation](#)

- Received: 30 November 2020
- Accepted: 06 April 2021
- Published: 12 May 2021
- Issue Date: 20 May 2021
- DOI: <https://doi.org/10.1038/s41586-021-03517-z>

## Comments

By submitting a comment you agree to abide by our [Terms](#) and [Community Guidelines](#). If you find something abusive or that does not comply with our terms or guidelines please flag it as inappropriate.

[Download PDF](#)

Advertisement

---

This article was downloaded by **calibre** from <https://www.nature.com/articles/s41586-021-03517-z>

- Article
- [Published: 21 April 2021](#)

# Structure of human telomerase holoenzyme with bound telomeric DNA

- [George E. Ghanim](#)<sup>1</sup> na1,
- [Adam J. Fountain](#)<sup>1</sup> na1,
- [Anne-Marie M. van Roon](#)<sup>1</sup> na1,
- [Ramya Rangan](#) ORCID: orcid.org/0000-0002-0960-0825<sup>2</sup>,
- [Rhiju Das](#) ORCID: orcid.org/0000-0001-7497-0972<sup>2,3,4</sup>,
- [Kathleen Collins](#)<sup>5,6</sup> &
- [Thi Hoang Duong Nguyen](#) ORCID: orcid.org/0000-0002-2180-3430<sup>1</sup>

[Nature](#) volume 593, pages 449–453 (2021) [Cite this article](#)

- 8361 Accesses
- 60 Altmetric
- [Metrics details](#)

## Subjects

- [Cryoelectron microscopy](#)
- [Enzyme mechanisms](#)
- [RNA](#)
- [Telomeres](#)

## Abstract

Telomerase adds telomeric repeats at chromosome ends to compensate for the telomere loss that is caused by incomplete genome end replication<sup>1</sup>. In humans, telomerase is upregulated during embryogenesis and in cancers, and mutations that compromise the function of telomerase result in disease<sup>2</sup>. A previous structure of human telomerase at a resolution of 8 Å revealed a vertebrate-specific composition and architecture<sup>3</sup>, comprising a catalytic core that is flexibly tethered to an H and ACA (hereafter, H/ACA) box ribonucleoprotein (RNP) lobe by telomerase RNA. High-resolution structural information is necessary to develop treatments that can effectively modulate telomerase activity as a therapeutic approach against cancers and disease. Here we used cryo-electron microscopy to determine the structure of human telomerase holoenzyme bound to telomeric DNA at sub-4 Å resolution, which reveals crucial DNA- and RNA-binding interfaces in the active site of telomerase as well as the locations of mutations that alter telomerase activity. We identified a histone H2A–H2B dimer within the holoenzyme that was bound to an essential telomerase RNA motif, which suggests a role for histones in the folding and function of telomerase RNA. Furthermore, this structure of a eukaryotic H/ACA RNP reveals the molecular recognition of conserved RNA and protein motifs, as well as interactions that are crucial for understanding the molecular pathology of many mutations that cause disease. Our findings provide the structural details of the assembly and active site of human telomerase, which paves the way for the development of therapeutic agents that target this enzyme.

## Access options

Subscribe to Journal

Get full journal access for 1 year

\$199.00

only \$3.90 per issue

## Subscribe

All prices are NET prices.

VAT will be added later in the checkout.

Tax calculation will be finalised during checkout.

Rent or Buy article

Get time limited or full article access on ReadCube.

from \$8.99

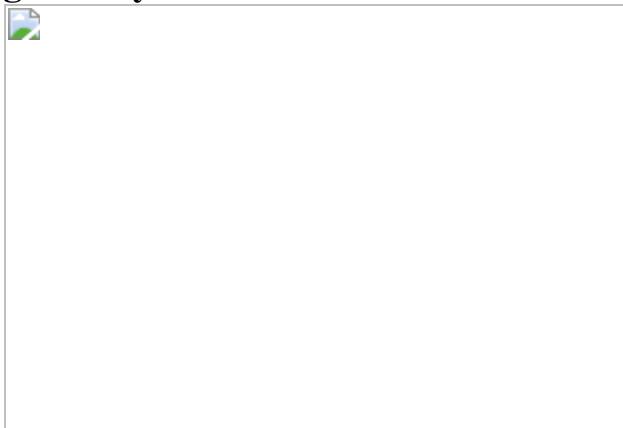
## Rent or Buy

All prices are NET prices.

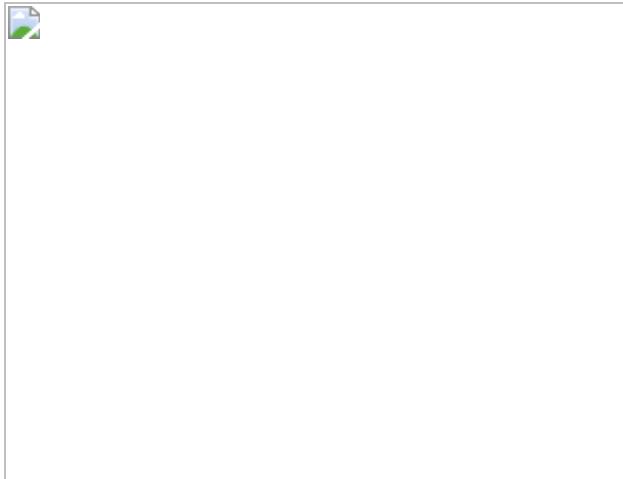
## **Additional access options:**

- [Log in](#)
- [Access through your institution](#)
- [Learn about institutional subscriptions](#)

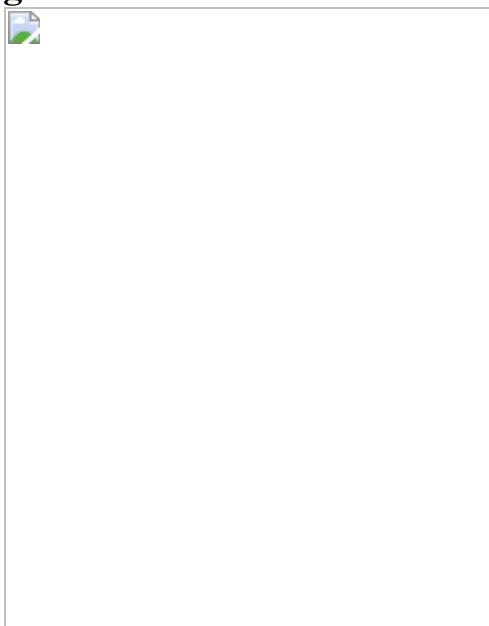
**Fig. 1: Cryo-EM structure of human telomerase holoenzyme.**



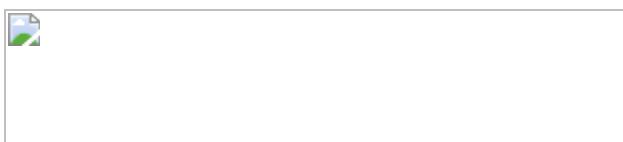
**Fig. 2: Catalytic core of human telomerase.**



**Fig. 3: CR4/5 interactions with TERT and histone H2A–H2B subunits.**



**Fig. 4: H/ACA RNP structure and mutations linked with disease in humans.**



## Data availability

Raw gels are provided in Supplementary Fig. 1. Replicates of the activity assays and immunoblotting experiments shown in Fig. 3d,e and quantification are included in Supplementary Fig. 2. Full mass spectrometry

data and coordinates of hTR from DRRAFTER modelling are provided in [the Supplementary Information](#). Cryo-EM maps of the catalytic core, H/ACA lobes and the whole telomerase maps (class 2 and class 5) with both lobes have been deposited with the Electron Microscopy Data Bank under accession numbers [EMD-12174](#), [EMD-12177](#), [EMD-12175](#) and [EMD-12176](#). Refined atomic coordinates for the catalytic core and H/ACA lobes are deposited with the PDB under accession numbers [7BG9](#) and [7BGB](#). The rigid body fitted models of the catalytic core and the H/ACA lobes into the whole telomerase maps (class 2 and class 5) are included as a Pymol session in Supplementary Data [2](#), [3](#). The BALBES database was provided internally by the BALBES developers, and is available upon request to the corresponding author. [Source data](#) are provided with this paper.

## References

1. 1.

Levy, M. Z., Allsopp, R. C., Futcher, A. B., Greider, C. W. & Harley, C. B. Telomere end-replication problem and cell aging. *J. Mol. Biol.* **225**, 951–960 (1992).

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

2. 2.

Shay, J. W. Role of telomeres and telomerase in aging and cancer. *Cancer Discov.* **6**, 584–593 (2016).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

3. 3.

Nguyen, T. H. D. et al. Cryo-EM structure of substrate-bound human telomerase holoenzyme. *Nature* **557**, 190–195 (2018).

[ADS](#) [CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

4. 4.

Blackburn, E. H. & Collins, K. Telomerase: an RNP enzyme synthesizes DNA. *Cold Spring Harb. Perspect. Biol.* **3**, a003558 (2011).

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

5. 5.

Lingner, J. et al. Reverse transcriptase motifs in the catalytic subunit of telomerase. *Science* **276**, 561–567 (1997).

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

6. 6.

Wu, R. A., Upton, H. E., Vogan, J. M. & Collins, K. Telomerase mechanism of telomere synthesis. *Annu. Rev. Biochem.* **86**, 439–460 (2017).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

7. 7.

MacNeil, D. E., Bensoussan, H. J. & Autexier, C. Telomerase regulation from beginning to the end. *Genes* **7**, 64 (2016).

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

8. 8.

Yu, Y.-T. & Meier, U. T. RNA-guided isomerization of uridine to pseudouridine—pseudouridylation. *RNA Biol.* **11**, 1483–1494 (2014).

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

9. 9.

Sarek, G., Marzec, P., Margalef, P. & Boulton, S. J. Molecular basis of telomere dysfunction in human genetic diseases. *Nat. Struct. Mol. Biol.* **22**, 867–874 (2015).

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

10. 10.

Li, L. & Ye, K. Crystal structure of an H/ACA box ribonucleoprotein particle. *Nature* **443**, 302–307 (2006).

[ADS](#) [CAS](#) [PubMed](#) [Google Scholar](#)

11. 11.

Li, S. et al. Reconstitution and structural analysis of the yeast box H/ACA RNA-guided pseudouridine synthase. *Genes Dev.* **25**, 2409–2421 (2011).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

12. 12.

Bai, X. C., Rajendra, E., Yang, G., Shi, Y. & Scheres, S. H. W. Sampling the conformational space of the catalytic subunit of human  $\gamma$ -secretase. *eLife* **4**, e11182 (2015).

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

13. 13.

Nakane, T., Kimanius, D., Lindahl, E. & Scheres, S. H. W. Characterisation of molecular motions in cryo-EM single-particle data by multi-body refinement in RELION. *eLife* **7**, e36861 (2018).

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

14. 14.

Kappel, K. et al. De novo computational RNA modeling into cryo-EM maps of large ribonucleoprotein complexes. *Nat. Methods* **15**, 947–954 (2018).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

15. 15.

Podlevsky, J. D. & Chen, J. J. Evolutionary perspectives of telomerase RNA structure and function. *RNA Biol.* **13**, 720–732 (2016).

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

16. 16.

Jiang, J. et al. Structure of telomerase with telomeric DNA. *Cell* **173**, 1179–1190 (2018).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

17. 17.

Wu, R. A., Tam, J. & Collins, K. DNA-binding determinants and cellular thresholds for human telomerase repeat addition processivity. *EMBO J.* **36**, 1908–1927 (2017).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

18. 18.

Qi, X. et al. RNA/DNA hybrid binding affinity determines telomerase template-translocation efficiency. *EMBO J.* **31**, 150–161 (2012).

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

19. 19.

Bryan, T. M., Goodrich, K. J. & Cech, T. R. A mutant of *Tetrahymena* telomerase reverse transcriptase with increased processivity. *J. Biol.*

*Chem.* **275**, 24199–24207 (2000).

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

20. 20.

Schaich, M. A. et al. Mechanisms of nucleotide selection by telomerase. *eLife* **9**, e55438 (2020).

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

21. 21.

Miller, M. C., Liu, J. K. & Collins, K. Template definition by *Tetrahymena* telomerase reverse transcriptase. *EMBO J.* **19**, 4412–4422 (2000).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

22. 22.

Nair, D. T., Johnson, R. E., Prakash, S., Prakash, L. & Aggarwal, A. K. Replication by human DNA polymerase- $\eta$  occurs by Hoogsteen base-pairing. *Nature* **430**, 377–380 (2004).

[ADS](#) [CAS](#) [PubMed](#) [Google Scholar](#)

23. 23.

Chen, J. L., Opperman, K. K. & Greider, C. W. A critical stem-loop structure in the CR4–CR5 domain of mammalian telomerase RNA. *Nucleic Acids Res.* **30**, 592–597 (2002).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

24. 24.

Robart, A. R. & Collins, K. Investigation of human telomerase holoenzyme assembly, activity, and processivity using disease-linked

subunit variants. *J. Biol. Chem.* **285**, 4375–4386 (2010).

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

25. 25.

Kim, N. K., Theimer, C. A., Mitchell, J. R., Collins, K. & Feigon, J. Effect of pseudouridylation on the structure and activity of the catalytically essential P6.1 hairpin in human telomerase RNA. *Nucleic Acids Res.* **38**, 6746–6756 (2010).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

26. 26.

Long, F., Vagin, A. A., Young, P. & Murshudov, G. N. BALBES: a molecular-replacement pipeline. *Acta Crystallogr. D* **64**, 125–132 (2008).

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

27. 27.

Davey, C. A., Sargent, D. F., Luger, K., Maeder, A. W. & Richmond, T. J. Solvent mediated interactions in the structure of the nucleosome core particle at 1.9 Å resolution. *J. Mol. Biol.* **319**, 1097–1113 (2002).

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

28. 28.

Schnapp, G., Rodi, H.-P., Rettig, W. J., Schnapp, A. & Damm, K. One-step affinity purification protocol for human telomerase. *Nucleic Acids Res.* **26**, 3311–3313 (1998).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

29. 29.

Palka, C., Forino, N. M., Hentschel, J., Das, R. & Stone, M. D. Folding heterogeneity in the essential human telomerase RNA three-way junction. *RNA* **26**, 1787–1800 (2020).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

30. 30.

Fu, D. & Collins, K. Distinct biogenesis pathways for human telomerase RNA and H/ACA small nucleolar RNAs. *Mol. Cell* **11**, 1361–1372 (2003).

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

31. 31.

Egan, E. D. & Collins, K. An enhanced H/ACA RNP assembly mechanism for human telomerase RNA. *Mol. Cell. Biol.* **32**, 2428–2439 (2012).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

32. 32.

Tycowski, K. T., Shu, M. D., Kukoyi, A. & Steitz, J. A. A conserved WD40 protein binds the Cajal body localization signal of scaRNP particles. *Mol. Cell* **34**, 47–57 (2009).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

33. 33.

Venteicher, A. S. et al. A human telomerase holoenzyme protein required for Cajal body localization and telomere synthesis. *Science* **323**, 644–648 (2009).

[ADS](#) [CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

34. 34.

Theimer, C. A. et al. Structural and functional characterization of human telomerase RNA processing and cajal body localization signals. *Mol. Cell* **27**, 869–881 (2007).

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

35. 35.

Jády, B. E., Bertrand, E. & Kiss, T. Human telomerase RNA and box H/ACA scaRNAs share a common Cajal body-specific localization signal. *J. Cell Biol.* **164**, 647–652 (2004).

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

36. 36.

Parks, J. W. & Stone, M. D. Coordinated DNA dynamics during the human telomerase catalytic cycle. *Nat. Commun.* **5**, 4146 (2014).

[ADS](#) [CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

37. 37.

Patrick, E. M., Slivka, J. D., Payne, B., Comstock, M. J. & Schmidt, J. C. Observation of processive telomerase catalysis using high-resolution optical tweezers. *Nat. Chem. Biol.* **16**, 801–809 (2020).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

38. 38.

Nandakumar, J. et al. The TEL patch of telomere protein TPP1 mediates telomerase recruitment and processivity. *Nature* **492**, 285–289 (2012).

[ADS](#) [CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

39. 39.

Latrick, C. M. & Cech, T. R. POT1–TPP1 enhances telomerase processivity by slowing primer dissociation and aiding translocation. *EMBO J.* **29**, 924–933 (2010).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

40. 40.

Zhong, F. L. et al. TPP1 OB-fold domain controls telomere maintenance by recruiting telomerase to chromosome ends. *Cell* **150**, 481–494 (2012).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

41. 41.

Sexton, A. N., Youmans, D. T. & Collins, K. Specificity requirements for human telomere protein interaction with telomerase holoenzyme. *J. Biol. Chem.* **287**, 34455–34464 (2012).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

42. 42.

Upton, H. E., Hong, K. & Collins, K. Direct single-stranded DNA binding by Teb1 mediates the recruitment of *Tetrahymena thermophila* telomerase to telomeres. *Mol. Cell. Biol.* **34**, 4200–4212 (2014).

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

43. 43.

Jiang, J. et al. Structure of *Tetrahymena* telomerase reveals previously unknown subunits, functions, and interactions. *Science* **350**, aab4070 (2015).

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

44. 44.

Witkin, K. L. & Collins, K. Holoenzyme proteins required for the physiological assembly and activity of telomerase. *Genes Dev.* **18**, 1107–1118 (2004).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

45. 45.

Stone, M. D. et al. Stepwise protein-mediated RNA folding directs assembly of telomerase ribonucleoprotein. *Nature* **446**, 458–461 (2007).

[ADS](#) [CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

46. 46.

Greider, C. W. Regulating telomere length from the inside out: the replication fork model. *Genes Dev.* **30**, 1483–1491 (2016).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

47. 47.

Margalef, P. et al. Stabilization of reversed replication forks by telomerase drives telomere catastrophe. *Cell* **172**, 439–453 (2018).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

48. 48.

Sauer, P. V. et al. Mechanistic insights into histone deposition and nucleosome assembly by the chromatin assembly factor-1. *Nucleic Acids Res.* **46**, 9907–9917 (2018).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

49. 49.

Wu, R. A., Dagdas, Y. S., Yilmaz, S. T., Yildiz, A. & Collins, K. Single-molecule imaging of telomerase reverse transcriptase in human telomerase holoenzyme and minimal RNP complexes. *eLife* **4**, e08363 (2015).

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

50. 50.

Luger, K., Rechsteiner, T. J. & Richmond, T. J. in *Methods in Enzymology* Vol. 304 (eds. Wassarman P. M. & Wolffe, A. P.) 3–19 (Elsevier, 1999).

51. 51.

Nguyen, T. H. D. *Structural and Biochemical Studies of Spliceosomal Activation*. PhD thesis, Univ. of Cambridge (2014).

52. 52.

Goodrich, J. A. & Kugel, J. F. *Binding and Kinetics for Molecular Biologists* (Cold Spring Harbor Laboratory, 2007).

53. 53.

Mastronarde, D. N. Automated electron microscope tomography using robust prediction of specimen movements. *J. Struct. Biol.* **152**, 36–51 (2005).

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

54. 54.

Zivanov, J. et al. New tools for automated high-resolution cryo-EM structure determination in RELION-3. *eLife* **7**, e42166 (2018).

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

55. 55.

Rohou, A. & Grigorieff, N. CTFFIND4: fast and accurate defocus estimation from electron micrographs. *J. Struct. Biol.* **192**, 216–221 (2015).

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

56. 56.

Zivanov, J., Nakane, T. & Scheres, S. H. W. Estimation of high-order aberrations and anisotropic magnification from cryo-EM data sets in RELION-3.1. *IUCrJ* **7**, 253–267 (2020).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

57. 57.

Rosenthal, P. B. & Henderson, R. Optimal determination of particle orientation, absolute hand, and contrast loss in single-particle electron cryomicroscopy. *J. Mol. Biol.* **333**, 721–745 (2003).

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

58. 58.

Chen, S. et al. High-resolution noise substitution to measure overfitting and validate resolution in 3D structure determination by single particle electron cryomicroscopy. *Ultramicroscopy* **135**, 24–35 (2013).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

59. 59.

Brown, A. et al. Tools for macromolecular model building and refinement into electron cryo-microscopy reconstructions. *Acta Crystallogr. D* **71**, 136–153 (2015).

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

60. 60.

Vagin, A. & Teplyakov, A. Molecular replacement with MOLREP. *Acta Crystallogr. D* **66**, 22–25 (2010).

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

61. 61.

Casañal, A., Lohkamp, B. & Emsley, P. Current developments in *Coot* for macromolecular model building of electron cryo-microscopy and crystallographic data. *Protein Sci.* **29**, 1055–1064 (2020).

[Google Scholar](#)

62. 62.

Chou, F.-C., Echols, N., Terwilliger, T. C. & Das, R. RNA structure refinement using the ERRASER–Phenix pipeline. *Methods Mol. Biol.* **1320**, 269–282 (2016).

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

63. 63.

Afonine, P. V. et al. Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallogr. D* **68**, 352–367 (2012).

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

64. 64.

Murshudov, G. N. et al. REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr. D* **67**, 355–367 (2011).

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

65. 65.

Nicholls, R. A., Fischer, M., McNicholas, S. & Murshudov, G. N. Conformation-independent structural comparison of macromolecules with ProSMART. *Acta Crystallogr. D* **70**, 2487–2499 (2014).

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

66. 66.

Williams, C. J. et al. MolProbity: more and better reference data for improved all-atom structure validation. *Protein Sci.* **27**, 293–315 (2018).

[CAS](#) [Google Scholar](#)

67. 67.

Tang, G. et al. EMAN2: an extensible image processing suite for electron microscopy. *J. Struct. Biol.* **157**, 38–46 (2007).

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

68. 68.

Goddard, T. D. et al. UCSF ChimeraX: meeting modern challenges in visualization and analysis. *Protein Sci.* **27**, 14–25 (2018).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

69. 69.

Pettersen, E. F. et al. UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

70. 70.

Podlevsky, J. D., Bley, C. J., Omana, R. V., Qi, X. & Chen, J. J. The telomerase database. *Nucleic Acids Res.* **36**, D339–D343 (2008).

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

71. 71.

Tesmer, V. M., Smith, E. M., Danciu, O., Padmanaban, S. & Nandakumar, J. Combining conservation and species-specific differences to determine how human telomerase binds telomeres. *Proc. Natl Acad. Sci. USA* **116**, 26505–26515 (2019).

[CAS](#) [Google Scholar](#)

[Download references](#)

## Acknowledgements

We thank the MRC-LMB EM facility staff for access and support of electron microscopy sample preparation and data collection; J. Grimmett and T. Darling for maintaining the computing facility; the LMB mass spectrometry facility for running mass spectrometry experiments; S. Scotcher and the LMB workshop team for making our electrophoresis gel systems; K. Muir and D. Barford for sharing human histone expression constructs; K. Kappel for advice on DRRAFTER modelling; the K. Nagai, J. Löwe and L. Passmore laboratories for sharing reagents and equipment; members of the laboratories of K.C. and E. Nogales for past technical support; and D. Barford, A. Carter, E. Nogales, L. Passmore and S. Scheres for critical reading of the manuscript. R.R. is supported by a National Science Foundation Graduate Fellowship. This work was funded by a UKRI-Medical Research Council grant to T.H.D.N. (MC\_UP\_1201/19), an NIH grant to K.C. (GM054198) and NIH grants to R.D. (GM122579 and GM121487).

## Author information

Author notes

1. These authors contributed equally: George E. Ghanim, Adam J. Fountain, Anne-Marie M. van Roon

## Affiliations

1. Medical Research Council Laboratory of Molecular Biology, Cambridge, UK

George E. Ghanim, Adam J. Fountain, Anne-Marie M. van Roon & Thi Hoang Duong Nguyen

2. Biophysics Program, Stanford University, Stanford, CA, USA

Ramya Rangan & Rhiju Das

3. Department of Biochemistry, Stanford University, Stanford, CA, USA

Rhiju Das

4. Department of Physics, Stanford University, Stanford, CA, USA

Rhiju Das

5. Department of Molecular and Cell Biology, University of California, Berkeley, CA, USA

Kathleen Collins

6. California Institute for Quantitative Biology (QB3), University of California, Berkeley, CA, USA

Kathleen Collins

## Authors

1. George E. Ghanim

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

2. Adam J. Fountain

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

3. Anne-Marie M. van Roon

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

4. Ramya Rangan

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

5. Rhiju Das

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

6. Kathleen Collins

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

7. Thi Hoang Duong Nguyen

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

## Contributions

K.C. and T.H.D.N. initiated the project. T.H.D.N. collected and analysed electron microscopy data. G.E.G., A.J.F., A.-M.M.v.R. and T.H.D.N. performed manual model building and refinement, and analysed the structures. R.R. and R.D. performed all DRRAFTER modelling of RNA and ERRASER for improving RNA geometry. G.E.G., A.-M.M.v.R. and T.H.D.N. performed all biochemical experiments. G.E.G. and A.J.F.

performed all quantifications. T.H.D.N. wrote the paper with inputs from all authors.

## Corresponding author

Correspondence to [Thi Hoang Duong Nguyen](#).

## Ethics declarations

### Competing interests

The authors declare no competing interests.

## Additional information

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

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

## Extended data figures and tables

### [Extended Data Fig. 1 Image processing scheme.](#)

Summary of the data processing strategies that yielded the reconstructions described in this Article.

### [Extended Data Fig. 2 Local resolution and resolution estimation.](#)

**a**, Gold-standard FSC curves for maps of H/ACA lobe, the catalytic core and the best two full-length classes. Resolutions were estimated at FSC = 0.143. **b**, FSC curves for the refined H/ACA and catalytic core

models versus the corresponding maps. Resolutions were estimated at FSC = 0.5. **c–f**, Local resolution estimated by RELION 3.1<sup>54</sup> for the H/ACA lobe (**c**), the catalytic core (**d**), full-length 3D class 2 (**e**) and full-length 3D class 5 (**f**). For direct comparisons, the same local resolution range (3.3–16 Å) was used for all maps. [Source data](#)

### Extended Data Fig. 3 Electron microscopy density of protein components.

**a, e, f, i, k, l**, Full densities of TERT (**a**), histone H2A (**e**), histone H2B (**f**), TCAB1 (**i**), 5' H/ACA tetramer (**k**) and 3' H/ACA tetramer (**l**). **b**, Close-up view of the active site density of TERT with an empty nucleotide-binding pocket (Extended Data Fig. [6f](#)). **c, d**, Representative electron microscopy densities of TERT. **g, h, j**, Representative electron microscopy densities of histone H2A (**g**), histone H2B (**h**) and TCAB1 (**j**). **m**, Close-up view of the density of the N-terminal extension of the 5' dyskerin bound to a hydrophobic pocket within the 3' dyskerin (Fig. [4c](#)). **n**, Close-up view of the density of helix 352–375 of the 3' dyskerin bound to the equivalent hydrophobic pocket as in **m** within the 5' dyskerin (Fig. [4d](#)). **o**, Close-up view of the density of the P8 stem loop of hTR, which contains the CAB and BIO boxes that interact with TCAB1 and the 3' NHP2 (Extended Data Figs. [4l](#), [8d](#)). **p**, Close-up view of the density of the H and ACA boxes interacting with the two dyskerin molecules (Extended Data Figs. [4k](#), [8c](#)).

### Extended Data Fig. 4 Electron microscopy density of hTR and DNA substrate.

**a**, Secondary structure schematic of hTR based on the structure. This figure was modified from the telomerase database<sup>70</sup>. **b, c**, Full density of hTR in the catalytic core (**b**) and the H/ACA lobe (**c**). **d**, Close-up view of the density of interactions between the DNA substrate and hTR template (Fig. [2b](#)). **e**, Close-up view of density of the DNA substrate and neighbouring residues of TERT (Extended Data Fig. [6d](#)). **f**, Close-up view of density of the RNA template region and neighbouring residues of TERT (Extended Data Fig. [6e](#)). **g**, Density of the P6.1 hairpin of the CR4/5 domain. Labelled residues are highlighted in Fig. [3a](#). **h**, Close-up review of the density of the

P6.1 stem loop interacting with residues of the CTE domain of TERT (**g**, Fig. 3a). **i**, Close-up view of the density of residue L1019 of TERT, which interacts with the two flipped-out nucleotides, U177 of the PK and U307 of the P6.1 stem loop as highlighted in Fig. 3b. **j**, Representative density of the PK containing the base triples, which are highlighted in black. Nucleotide U113 is modelled but not visible in this view. **k**, Density of the H and ACA boxes (Extended Data Fig. 3p). **l**, Density of the P8 stem loop (Extended Data Fig. 3o).

### Extended Data Fig. 5 Multibody refinement and conformational dynamics analysis of the full-length structure.

**a**, Summary of the multibody refinement strategy and principal component analysis<sup>13</sup>. The best two subsets from global 3D classification were subjected to multibody refinement using two masks for the H/ACA lobe (yellow) and the catalytic core (cyan). **b**, Principal component analysis for 3D class 2. **c**, Principal component analysis for 3D class 5. The first and last frames of the eigenvector series of the first six principal components are shown. Curved arrows indicate the movements. **d**, Top 10 hTR ensembles modelled by DRRAFTER into the refined full-length 3D class 2 map (Supplementary Data 2). **e**, Top 10 hTR ensembles modelled by DRRAFTER into the refined full-length 3D class 5 map (Supplementary Data 3). hTR is shown in blue, and the protein subunits are in grey.

### Extended Data Fig. 6 Telomerase catalytic cycle and DNA path.

**a**, Domain architecture of human TERT and conserved motifs that are often observed in reverse transcriptases<sup>70</sup>. **b**, TERT conserved motifs shown in **a**, and their interactions with the template and DNA substrate. **c**, A model for the telomerase catalytic cycle. The template region of hTR is divided into an alignment region and a template region. The telomeric DNA repeat first binds to the alignment region, followed by six consecutive nucleotide additions using the template region. After the synthesis of the full telomeric repeat, the DNA substrate translocates to bind the alignment region to start another round of repeat synthesis. The state captured in our structure is indicated with an asterisk. **d**, Interactions between the 3' telomeric

TTAGGG repeat of the substrate and TERT. **e**, Interactions between the template region of hTR and TERT. **f**, TERT active site in a prenucleotide state. D712, D868 and D869 form the catalytic triad for nucleotide addition (Extended Data Fig. 3b). **g**, Modelled dTTP (PDB 1T3N<sup>22</sup>) in the vacant nucleotide site of TERT. The C2 ribose is indicated. **h**, Structure of human TERT with PK/t and CR4/5 domains of hTR and DNA. **i**, Electrostatic surface potential of human TERT with hTR and the DNA substrate shown in the same view as in **h**. The highlighted human TEN–IFD–TRAP interface (in blue) is positively charged and could potentially bind the 5' end of the DNA substrate in human telomerase. **j**, IFD–TRAP and TEN domains of TERT. Residues that are known to affect TPP1 binding to TERT are highlighted as spheres<sup>71</sup>. The proposed DNA path would bring it close to the proposed TPP1-binding site on the TEN domain. **k**, Model of TPP1–POT1 bound to human TEN domain, based on the *Tetrahymena* p50–TEB complex. Despite the similar overall domain arrangements, the domains of *Tetrahymena* and human TERT do not align well as a whole. To obtain the model, we superposed the *Tetrahymena* TEN domain–p50–TEB1–TEB2–TEB3 complex (PDB 6D6V<sup>16</sup>) onto the human TEN domain. **l**, Model of telomerase catalytic cycle. Telomerase template RNA binds the telomeric DNA substrate by base-pairing. The DNA-binding sites on TERT are indicated by yellow stars. One binding site is provided by motif T and the CTE domain of TERT near the active site, as observed in the structure (Fig. 2d and d). The second binding site is proposed to be provided by the TEN domain at the 5' end of the DNA. After the synthesis of a full telomeric repeat, the nascent DNA undergoes translocation and rebinding to the template RNA. We propose that the two DNA-binding sites form an anchor site to allow DNA retention for multiple rounds of repeat synthesis.

### **Extended Data Fig. 7 Identification of histone H2A and H2B as subunits of the human telomerase holoenzyme.**

- a**, Three-dimensional classification of the catalytic core (dataset 1), showing the presence of the unaccounted-for density (in yellow). The best class (boxed) has the most homogenous density, and was selected for the final refinement. Similar observations were made with the second dataset.
- b**, The 8 Å catalytic core map (in grey) with the previously unmodelled

density (in yellow)<sup>3</sup> (left) and with the model of the catalytic core obtained from this work fitted into it (right). The density assigned as the histone H2A–H2B dimer coincided with the unmodelled density from the previous work. **c**, The refined catalytic core map, with hTR, TERT and histone H2A–H2B segmented in different colours. **d**, Top, interactions between CR4/5 and the histone H2A–H2B dimer in human telomerase. Bottom, electrostatic surface potential of the histone dimer and the positively charged surface used for interacting with the CR4/5. **e**, Nucleosome structure. The histone H2A–H2B dimer is coloured and oriented the same way as in **d**. This shows that the histone H2A–H2B dimer uses the same surface to bind both nucleosomal DNA and CR4/5 (PDB 1KX5<sup>27</sup>). **f**, Purified histone H2A–H2B and CR4/5 RNA used for electrophoretic mobility shift assay in Fig. [3f](#). No RNA ladders were loaded with the CR4/5 RNA. **g**, Immunoblots of crude lysate of HEK 293T cells transfected with twin Strep-TERT and hTR expression constructs (input) and oligonucleotide elution from 2'-*O*-methyl purification (Fig. [3c](#)). These samples were immunoblotted for Strep and histones H2A, H2B, H3 and H4. The presence of histones H2A, H2B, H3 and H4 was also confirmed by mass spectrometry (Supplementary Data 4). **h**, Immunoblots of crude lysate of HEK 293T cells (input) and oligonucleotide elution from 2'-*O*-methyl purification. These samples were immunoblotted for dyskerin, and histones H2A, H2B, H3 and H4. **i**, Structure of the *Tetrahymena* TRBD–CTE–p65 and stem loop 4 (SL4) (PDB 6D6V<sup>16</sup>). **j**, Structure of human TRBD–CTE–H2A–H2B and CR4/5. **k**, Quantification of the electrophoretic mobility shift assays shown in Fig. [3f](#) for  $K_d$  determination. ‘Total fraction bound’ reflects the quantification of free probe depletion against total probe with increasing histone concentration. ‘Specific fraction bound’ reflects the quantification of the increasing discrete shifted complex band against total probe with increasing histone concentration. Points represent values from three independent replicates (Supplementary Fig. [1](#)). **l**, Superposition of the histone octamer structure, with flexible histone tails removed, onto the histone H2A–H2B dimer bound to the human telomerase catalytic core in two different views (PDB 1KX5<sup>27</sup>).

## Extended Data Fig. 8 H/ACA RNP and molecular interactions of conserved RNA motifs.

**a**, Front (left) and back (right) views of the H/ACA RNP with subunits coloured as indicated. **b**, Secondary structure schematic of the H/ACA domain of hTR. **c**, Left, close-up view of the H and ACA boxes, and their interactions with each other and with the two dyskerin molecules (Extended Data Figs. 3p, 4k). Right, sequences of H and ACA boxes, with conserved nucleotides highlighted in bold. **d**, Left, close-up view of P8 stem loop and interactions between the CAB and BIO boxes with TCAB1 and NHP2 (Extended Data Figs. 3o, 4l). Right, a schematic of these interactions.

**Extended Data Table 1 Cryo-EM data collection, refinement and validation statistics**

[Full size table](#)

**Extended Data Table 2 Modelling of protein and RNA components**

[Full size table](#)

## Supplementary information

### [Supplementary Information](#)

This file contains Supplementary Text, Supplementary Figures 1 – 2 and Supplementary Table 1.

### [Reporting Summary](#)

### [Supplementary Data 1](#)

Pymol session showing the top 10 hTR models obtained from DRRAFTER for the catalytic core.

### [Supplementary Data 2](#)

### [Supplementary Data 3](#)

### [Supplementary Data 4](#)

### [Peer Review File](#)

## Source data

[Source Data Extended Data Fig. 2](#)

## Rights and permissions

[Reprints and Permissions](#)

## About this article



Check for  
updates

## Cite this article

Ghanim, G.E., Fountain, A.J., van Roon, AM.M. *et al.* Structure of human telomerase holoenzyme with bound telomeric DNA. *Nature* **593**, 449–453 (2021). <https://doi.org/10.1038/s41586-021-03415-4>

[Download citation](#)

- Received: 30 September 2020
- Accepted: 03 March 2021
- Published: 21 April 2021
- Issue Date: 20 May 2021
- DOI: <https://doi.org/10.1038/s41586-021-03415-4>

## Comments

By submitting a comment you agree to abide by our [Terms](#) and [Community Guidelines](#). If you find something abusive or that does not comply with our terms or guidelines please flag it as inappropriate.

[Access through your institution](#)

[Change institution](#)

[Buy or subscribe](#)

Advertisement

---

This article was downloaded by **calibre** from <https://www.nature.com/articles/s41586-021-03415-4>

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

- Article
- [Published: 12 May 2021](#)

# Structures of telomerase at several steps of telomere repeat synthesis

- [Yao He](#) [ORCID: orcid.org/0000-0002-1368-6422<sup>1,2</sup>](#),
- [Yaqiang Wang](#) [ORCID: orcid.org/0000-0002-6822-4778<sup>1</sup>](#),
- [Baocheng Liu<sup>1</sup>](#),
- [Christina Helmling<sup>1</sup>](#),
- [Lukas Sušac<sup>1</sup>](#),
- [Ryan Cheng<sup>1</sup>](#),
- [Z. Hong Zhou](#) [ORCID: orcid.org/0000-0002-8373-4717<sup>2,3</sup>](#) &
- [Juli Feigon](#) [ORCID: orcid.org/0000-0003-3376-435X<sup>1</sup>](#)

[Nature](#) volume **593**, pages 454–459 (2021) [Cite this article](#)

- 3620 Accesses
- 38 Altmetric
- [Metrics details](#)

## Subjects

- [Cryoelectron microscopy](#)
- [DNA synthesis](#)
- [Long non-coding RNAs](#)
- [RNA](#)
- [Telomeres](#)

## Abstract

Telomerase is unique among the reverse transcriptases in containing a noncoding RNA (known as telomerase RNA (TER)) that includes a short template that is used for the processive synthesis of G-rich telomeric DNA repeats at the 3' ends of most eukaryotic chromosomes<sup>1</sup>. Telomerase maintains genomic integrity, and its activity or dysregulation are critical determinants of human longevity, stem cell renewal and cancer progression<sup>2,3</sup>. Previous cryo-electron microscopy structures have established the general architecture, protein components and stoichiometries of *Tetrahymena* and human telomerase, but our understandings of the details of DNA–protein and RNA–protein interactions and of the mechanisms and recruitment involved remain limited<sup>4,5,6</sup>. Here we report cryo-electron microscopy structures of active *Tetrahymena* telomerase with telomeric DNA at different steps of nucleotide addition. Interactions between telomerase reverse transcriptase (TERT), TER and DNA reveal the structural basis of the determination of the 5' and 3' template boundaries, handling of the template–DNA duplex and separation of the product strand during nucleotide addition. The structure and binding interface between TERT and telomerase protein p50 (a homologue of human TPP1<sup>7,8</sup>) define conserved interactions that are required for telomerase activation and recruitment to telomeres. Telomerase La-related protein p65 remodels several regions of TER, bridging the 5' and 3' ends and the conserved pseudoknot to facilitate assembly of the TERT–TER catalytic core.

## Access options

Subscribe to Journal

Get full journal access for 1 year

\$199.00

only \$3.90 per issue

[Subscribe](#)

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

Rent or Buy article

Get time limited or full article access on ReadCube.

from \$8.99

[Rent or Buy](#)

All prices are NET prices.

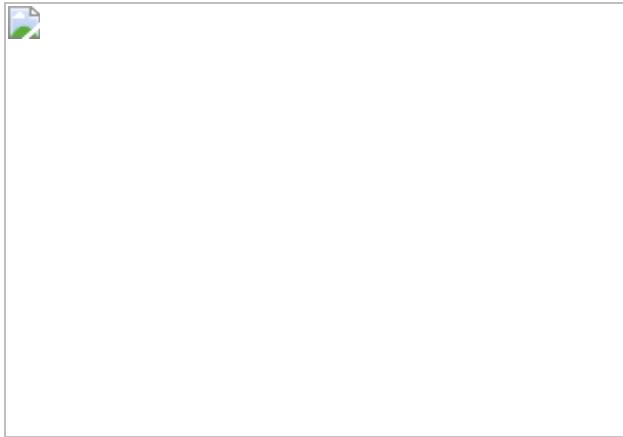
### **Additional access options:**

- [Log in](#)
- [Access through your institution](#)
- [Learn about institutional subscriptions](#)

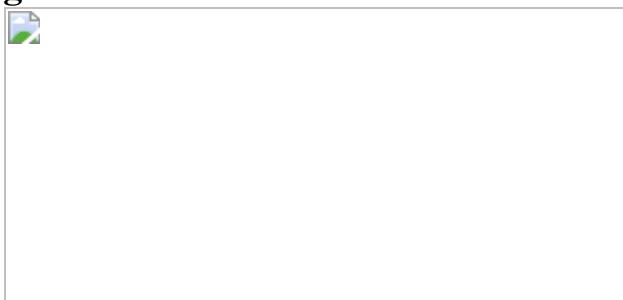
**Fig. 1: Structure of *Tetrahymena* telomerase with sstDNA.**



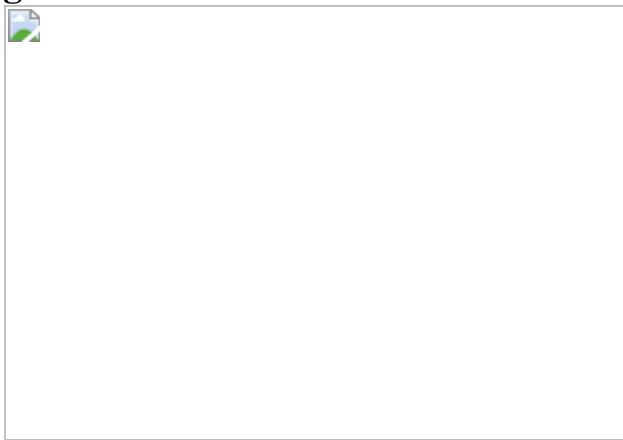
**Fig. 2: Structure of p50 OB and interactions with TERT.**



**Fig. 3: Interactions between TERT and template–DNA duplex.**



**Fig. 4: Structural details for determining template boundaries.**



**Fig. 5: A model for telomere repeat synthesis.**

## Data availability

Cryo-EM density maps have been deposited in the EMDB under accession numbers [EMD-23437](#) (telomerase T3D2), [EMD-23438](#) (telomerase T4D4) and [EMD-23439](#) (telomerase T5D5). The atomic models have been deposited in the PDB under accession codes [7LMA](#) (telomerase T3D2) and [7LMB](#) (telomerase T5D5). The atomic model and cryo-EM density map of telomerase T3D3 were retrieved from the PDB (accession code 6D6V) and EMDB (accession code EMD-7821). Other structures used in this study were retrieved from the PDB with accession codes 2I46 (TPP1 OB), 3KYL (*Tribolium* TERT-like protein) and 2M22 (TER stem-loop 2). Uncropped version of all the gels are included as Supplementary Fig. 1. Any other relevant data are available from the corresponding authors upon reasonable request.

## References

1. 1.

Blackburn, E. H. & Collins, K. Telomerase: an RNP enzyme synthesizes DNA. *Cold Spring Harb. Perspect. Biol.* **3**, a003558 (2011).

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

2. 2.

Armanios, M. & Blackburn, E. H. The telomere syndromes. *Nat. Rev. Genet.* **13**, 693–704 (2012).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

3. 3.

Shay, J. W. Role of telomeres and telomerase in aging and cancer. *Cancer Discov.* **6**, 584–593 (2016).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

4. 4.

Jiang, J. et al. Structure of *Tetrahymena* telomerase reveals previously unknown subunits, functions, and interactions. *Science* **350**, aab4070 (2015).

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

5. 5.

Jiang, J. et al. Structure of telomerase with telomeric DNA. *Cell* **173**, 1179–1190.e13 (2018).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

6. 6.

Nguyen, T. H. D. et al. Cryo-EM structure of substrate-bound human telomerase holoenzyme. *Nature* **557**, 190–195 (2018).

[ADS](#) [CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

7. 7.

Wang, F. et al. The POT1–TPP1 telomere complex is a telomerase processivity factor. *Nature* **445**, 506–510 (2007).

[ADS](#) [CAS](#) [PubMed](#) [Google Scholar](#)

8. 8.

Xin, H. et al. TPP1 is a homologue of ciliate TEBP- $\beta$  and interacts with POT1 to recruit telomerase. *Nature* **445**, 559–562 (2007).

[ADS](#) [CAS](#) [PubMed](#) [Google Scholar](#)

9. 9.

Greider, C. W. & Blackburn, E. H. Identification of a specific telomere terminal transferase activity in *Tetrahymena* extracts. *Cell* **43**, 405–413 (1985).

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

10. 10.

Chan, H., Wang, Y. & Feigon, J. Progress in human and *Tetrahymena* telomerase structure determination. *Annu. Rev. Biophys.* **46**, 199–225 (2017).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

11. 11.

Wu, R. A., Upton, H. E., Vogan, J. M. & Collins, K. Telomerase mechanism of telomere synthesis. *Annu. Rev. Biochem.* **86**, 439–460 (2017).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

12. 12.

Podlevsky, J. D., Bley, C. J., Omana, R. V., Qi, X. & Chen, J. J. The telomerase database. *Nucleic Acids Res.* **36**, D339–D343 (2008).

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

13. 13.

Prathapam, R., Witkin, K. L., O'Connor, C. M. & Collins, K. A telomerase holoenzyme protein enhances telomerase RNA assembly with telomerase reverse transcriptase. *Nat. Struct. Mol. Biol.* **12**, 252–257 (2005).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

14. 14.

Mitchell, J. R., Cheng, J. & Collins, K. A box H/ACA small nucleolar RNA-like domain at the human telomerase RNA 3' end. *Mol. Cell. Biol.* **19**, 567–576 (1999).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

15. 15.

Gillis, A. J., Schuller, A. P. & Skordalakes, E. Structure of the *Tribolium castaneum* telomerase catalytic subunit TERT. *Nature* **455**, 633–637 (2008).

[ADS](#) [CAS](#) [PubMed](#) [Google Scholar](#)

16. 16.

Jiang, J. et al. The architecture of *Tetrahymena* telomerase holoenzyme. *Nature* **496**, 187–192 (2013).

[ADS](#) [CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

17. 17.

Lei, M., Podell, E. R. & Cech, T. R. Structure of human POT1 bound to telomeric single-stranded DNA provides a model for chromosome end-protection. *Nat. Struct. Mol. Biol.* **11**, 1223–1229 (2004).

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

18. 18.

Lim, C. J. & Cech, T. R. Shaping human telomeres: from shelterin and CST complexes to telomeric chromatin organization. *Nat. Rev. Mol. Cell Biol.* **22**, 283–298 (2021).

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

19. 19.

Lue, N. F., Chan, J., Wright, W. E. & Hurwitz, J. The CDC13–STN1–TEN1 complex stimulates Pol  $\alpha$  activity by promoting RNA priming and primase-to-polymerase switch. *Nat. Commun.* **5**, 5762 (2014).

[ADS](#) [CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

20. 20.

Maraia, R. J., Mattijssen, S., Cruz-Gallardo, I. & Conte, M. R. The La and related RNA-binding proteins (LARPs): structures, functions, and evolving perspectives. *Wiley Interdiscip. Rev. RNA* **8**, e1430 (2017).

[Google Scholar](#)

21. 21.

Singh, M. et al. Structural basis for telomerase RNA recognition and RNP assembly by the holoenzyme La family protein p65. *Mol. Cell* **47**, 16–26 (2012).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

22. 22.

Stone, M. D. et al. Stepwise protein-mediated RNA folding directs assembly of telomerase ribonucleoprotein. *Nature* **446**, 458–461 (2007).

[ADS](#) [CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

23. 23.

Kotik-Kogan, O., Valentine, E. R., Sanfelice, D., Conte, M. R. & Curry, S. Structural analysis reveals conformational plasticity in the recognition of RNA 3' ends by the human La protein. *Structure* **16**, 852–862 (2008).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

24. 24.

Nandakumar, J. et al. The TEL patch of telomere protein TPP1 mediates telomerase recruitment and processivity. *Nature* **492**, 285–289 (2012).

[ADS](#) [CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

25. 25.

Zhong, F. L. et al. TPP1 OB-fold domain controls telomere maintenance by recruiting telomerase to chromosome ends. *Cell* **150**, 481–494 (2012).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

26. 26.

Tesmer, V. M., Smith, E. M., Danciu, O., Padmanaban, S. & Nandakumar, J. Combining conservation and species-specific differences to determine how human telomerase binds telomeres. *Proc. Natl Acad. Sci. USA* **116**, 26505–26515 (2019).

[CAS](#) [Google Scholar](#)

27. 27.

Grill, S., Tesmer, V. M. & Nandakumar, J. The N terminus of the OB domain of telomere protein TPP1 is critical for telomerase action. *Cell Rep.* **22**, 1132–1140 (2018).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

28. 28.

Zhang, Y. et al. Phosphorylation of TPP1 regulates cell cycle-dependent telomerase recruitment. *Proc. Natl Acad. Sci. USA* **110**, 5457–5462 (2013).

[ADS](#) [CAS](#) [PubMed](#) [Google Scholar](#)

29. 29.

Upton, H. E., Chan, H., Feigon, J. & Collins, K. Shared subunits of *Tetrahymena* telomerase holoenzyme and replication protein a have different functions in different cellular complexes. *J. Biol. Chem.* **292**, 217–228 (2017).

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

30. 30.

Zeng, Z. et al. Structural basis for *Tetrahymena* telomerase processivity factor Teb1 binding to single-stranded telomeric-repeat DNA. *Proc. Natl Acad. Sci. USA* **108**, 20357–20361 (2011).

[ADS](#) [CAS](#) [PubMed](#) [Google Scholar](#)

31. 31.

Shastry, S., Steinberg-Neifach, O., Lue, N. & Stone, M. D. Direct observation of nucleic acid binding dynamics by the telomerase essential N-terminal domain. *Nucleic Acids Res.* **46**, 3088–3102 (2018).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

32. 32.

Jansson, L. I. et al. Telomere DNA G-quadruplex folding within actively extending human telomerase. *Proc. Natl Acad. Sci. USA* **116**, 9350–9359 (2019).

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

33. 33.

Patrick, E. M., Slivka, J. D., Payne, B., Comstock, M. J. & Schmidt, J. C. Observation of processive telomerase catalysis using high-resolution optical tweezers. *Nat. Chem. Biol.* **16**, 801–809 (2020).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

34. 34.

Wang, Y., Sušac, L. & Feigon, J. Structural biology of telomerase. *Cold Spring Harb. Perspect. Biol.* **11**, a032383 (2019).

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

35. 35.

Xie, M., Podlevsky, J. D., Qi, X., Bley, C. J. & Chen, J. J. A novel motif in telomerase reverse transcriptase regulates telomere repeat addition rate and processivity. *Nucleic Acids Res.* **38**, 1982–1996 (2010).

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

36. 36.

Nakamura, T. M. et al. Telomerase catalytic subunit homologs from fission yeast and human. *Science* **277**, 955–959 (1997).

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

37. 37.

Jansson, L. I. et al. Structural basis of template-boundary definition in *Tetrahymena* telomerase. *Nat. Struct. Mol. Biol.* **22**, 883–888 (2015).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

38. 38.

Akiyama, B. M., Gomez, A. & Stone, M. D. A conserved motif in *Tetrahymena thermophila* telomerase reverse transcriptase is proximal to the RNA template and is essential for boundary definition. *J. Biol. Chem.* **288**, 22141–22149 (2013).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

39. 39.

Harkisheimer, M., Mason, M., Shuvaeva, E. & Skordalakes, E. A motif in the vertebrate telomerase N-terminal linker of TERT contributes to RNA binding and telomerase activity and processivity. *Structure* **21**, 1870–1878 (2013).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

40. 40.

Vester, B. & Wengel, J. LNA (locked nucleic acid): high-affinity targeting of complementary RNA and DNA. *Biochemistry* **43**, 13233–13241 (2004).

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

41. 41.

Schaich, M. A. et al. Mechanisms of nucleotide selection by telomerase. *eLife* **9**, e55438 (2020).

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

42. 42.

Lai, C. K., Miller, M. C. & Collins, K. Template boundary definition in *Tetrahymena* telomerase. *Genes Dev.* **16**, 415–420 (2002).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

43. 43.

Miller, M. C. & Collins, K. Telomerase recognizes its template by using an adjacent RNA motif. *Proc. Natl Acad. Sci. USA* **99**, 6585–6590 (2002).

[ADS](#) [CAS](#) [PubMed](#) [Google Scholar](#)

44. 44.

Berman, A. J., Akiyama, B. M., Stone, M. D. & Cech, T. R. The RNA accordion model for template positioning by telomerase RNA during telomeric DNA synthesis. *Nat. Struct. Mol. Biol.* **18**, 1371–1375 (2011).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

45. 45.

Wang, M. et al. Stringent control of the RNA-dependent RNA polymerase translocation revealed by multiple intermediate structures. *Nat. Commun.* **11**, 2605 (2020).

[ADS](#) [CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

46. 46.

Yang, W. & Lee, Y. S. A DNA-hairpin model for repeat-addition processivity in telomere synthesis. *Nat. Struct. Mol. Biol.* **22**, 844–847 (2015).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

47. 47.

Wu, R. A., Tam, J. & Collins, K. DNA-binding determinants and cellular thresholds for human telomerase repeat addition processivity. *EMBO J.* **36**, 1908–1927 (2017).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

48. 48.

Wang, Y., Gallagher-Jones, M., Sušac, L., Song, H. & Feigon, J. A structurally conserved human and *Tetrahymena* telomerase catalytic core. *Proc. Natl Acad. Sci. USA* **117**, 31078–31087 (2020).

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

49. 49.

Hu, X. et al. Quality-control mechanism for telomerase RNA folding in the cell. *Cell Rep.* **33**, 108568 (2020).

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

50. 50.

Holohan, B. et al. Impaired telomere maintenance in Alazami syndrome patients with LARP7 deficiency. *BMC Genomics* **17** (Suppl 9), 749 (2016).

[MathSciNet](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

51. 51.

Min, B. & Collins, K. An RPA-related sequence-specific DNA-binding subunit of telomerase holoenzyme is required for elongation processivity and telomere maintenance. *Mol. Cell* **36**, 609–619 (2009).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

52. 52.

Mastronarde, D. N. Automated electron microscope tomography using robust prediction of specimen movements. *J. Struct. Biol.* **152**, 36–51 (2005).

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

53. 53.

Zheng, S. Q. et al. MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat. Methods* **14**, 331–332 (2017).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

54. 54.

Rohou, A. & Grigorieff, N. CTFFIND4: fast and accurate defocus estimation from electron micrographs. *J. Struct. Biol.* **192**, 216–221 (2015).

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

55. 55.

Zivanov, J. et al. New tools for automated high-resolution cryo-EM structure determination in RELION-3. *eLife* **7**, e42166 (2018).

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

56. 56.

Rosenthal, P. B. & Henderson, R. Optimal determination of particle orientation, absolute hand, and contrast loss in single-particle electron cryomicroscopy. *J. Mol. Biol.* **333**, 721–745 (2003).

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

57. 57.

Kucukelbir, A., Sigworth, F. J. & Tagare, H. D. Quantifying the local resolution of cryo-EM density maps. *Nat. Methods* **11**, 63–65 (2014).

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

58. 58.

Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D* **66**, 486–501 (2010).

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

59. 59.

Pettersen, E. F. et al. UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

60. 60.

Buchan, D. W., Minneci, F., Nugent, T. C., Bryson, K. & Jones, D. T. Scalable web services for the PSIPRED protein analysis workbench. *Nucleic Acids Res.* **41**, W349–W357 (2013).

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

61. 61.

Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N. & Sternberg, M. J. The Phyre2 web portal for protein modeling, prediction and analysis. *Nat. Protoc.* **10**, 845–858 (2015).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

62. 62.

Richards, R. J., Theimer, C. A., Finger, L. D. & Feigon, J. Structure of the *Tetrahymena thermophila* telomerase RNA helix II template boundary element. *Nucleic Acids Res.* **34**, 816–825 (2006).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

63. 63.

Adams, P. D. et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D* **66**, 213–221 (2010).

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

64. 64.

Goddard, T. D. et al. UCSF ChimeraX: meeting modern challenges in visualization and analysis. *Protein Sci.* **27**, 14–25 (2018).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

65. 65.

Cash, D. D. & Feigon, J. Structure and folding of the *Tetrahymena* telomerase RNA pseudoknot. *Nucleic Acids Res.* **45**, 482–495 (2017).

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

66. 66.

Petrov, A., Wu, T., Puglisi, E. V. & Puglisi, J. D. RNA purification by preparative polyacrylamide gel electrophoresis. *Methods Enzymol.* **530**, 315–330 (2013).

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

67. 67.

Hong, K. et al. *Tetrahymena* telomerase holoenzyme assembly, activation, and inhibition by domains of the p50 central hub. *Mol. Cell. Biol.* **33**, 3962–3971 (2013).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

68. 68.

Latrick, C. M. & Cech, T. R. POT1-TPP1 enhances telomerase processivity by slowing primer dissociation and aiding translocation. *EMBO J.* **29**, 924–933 (2010).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

69. 69.

Chu, T. W., D’Souza, Y. & Autexier, C. The insertion in fingers domain in human telomerase can mediate enzyme processivity and telomerase recruitment to telomeres in a TPP1-dependent manner. *Mol. Cell. Biol.* **36**, 210–222 (2015).

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

70. 70.

Papadopoulos, J. S. & Agarwala, R. COBALT: constraint-based alignment tool for multiple protein sequences. *Bioinformatics* **23**, 1073–1079 (2007).

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

71. 71.

Jacobs, S. A., Podell, E. R. & Cech, T. R. Crystal structure of the essential N-terminal domain of telomerase reverse transcriptase. *Nat. Struct. Mol. Biol.* **13**, 218–225 (2006).

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

[Download references](#)

## Acknowledgements

This work was supported by NIH R35GM131901 and NSF MCB2016540 grants to J.F. and NIH grant GM071940 to Z.H.Z. We acknowledge use of instruments at the Electron Imaging Center for Nanomachines supported by

UCLA and by instrumentation grants from NIH (1S10RR23057, 1S10OD018111 and U24GM116792) and NSF (DBI-1338135 and DMR-1548924). Some preliminary data were collected at the Stanford-SLAC Cryo-EM Center (S2C2) supported by the NIH Common Fund Transformative High Resolution Cryo-Electron Microscopy programme (U24 GM129541) and National Center for CryoEM Access and Training (NCCAT) and the Simons Electron Microscopy Center located at the New York Structural Biology Center, supported by the NIH Common Fund Transformative High Resolution Cryo-Electron Microscopy programme (U24 GM129539), and by grants from the Simons Foundation (SF349247) and NY State Assembly. We thank D. Weisman for help with illustration of Fig. 5.

## Author information

### Affiliations

1. Department of Chemistry and Biochemistry, University of California, Los Angeles, Los Angeles, CA, USA

Yao He, Yaqiang Wang, Baocheng Liu, Christina Helmling, Lukas Sušac, Ryan Cheng & Juli Feigon

2. Department of Microbiology, Immunology, and Molecular Genetics, University of California, Los Angeles, Los Angeles, CA, USA

Yao He & Z. Hong Zhou

3. California NanoSystems Institute, University of California, Los Angeles, Los Angeles, CA, USA

Z. Hong Zhou

### Authors

1. Yao He  
[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

2. Yaqiang Wang

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

3. Baocheng Liu

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

4. Christina Helmling

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

5. Lukas Sušac

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

6. Ryan Cheng

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

7. Z. Hong Zhou

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

8. Juli Feigon

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

## Contributions

J.F. and Z.H.Z. supervised the project; Y.H. purified telomerase samples, screened cryo-EM grids, and performed cryo-EM data collection and processing; Y.H. and Y.W. built atomic models; Y.H. and B.L. performed telomerase activity assays; C.H., L.S. and R.C. helped with telomerase sample preparation; Y.H. and J.F. wrote the manuscript. All authors contributed to the final version.

## Corresponding authors

Correspondence to [Z. Hong Zhou](#) or [Juli Feigon](#).

## Ethics declarations

## Competing interests

The authors declare no competing interests.

## Additional information

**Peer review information** *Nature* thanks the anonymous reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

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

## Extended data figures and tables

### [Extended Data Fig. 1 Biochemical and biophysical evaluation of endogenously purified \*Tetrahymena\* telomerase with sstDNA.](#)

**a**, Silver-stained SDS-PAGE gel of the tandem-affinity-purified telomerase. Serial diluted BSA samples were loaded together to assist concentration estimation of the telomerase sample. Gel image is representative of independent biological replicates ( $n = 3$ ). **b**, Direct telomeric DNA

extension assays of the purified telomerase bound with different sstDNA primers. A standard telomere addition pattern is observed when using a (GTTGGG)<sub>5</sub> or (GTTGGG)<sub>3</sub> primer (P1 and P2). However, the translocation of product is inhibited when using (GTTGGG)<sub>2</sub>GTTGG<sup>L</sup>G<sup>L</sup>G<sup>L</sup>T primer (P3), resulting in a single dark band (red asterisk). G<sup>L</sup> denotes an LNA nucleotide instead of a DNA nucleotide. The LNA-containing product (red asterisk) migrates slightly slower through the gel as compared to nonmodified DNA. Gel image is representative of independent biological replicates ( $n = 3$ ). **c**, Motion-corrected cryo-EM micrograph. **d**, Representative 2D class averages of telomerase particles. All results from sample purification (**a**), activity assays (**b**) and cryo-EM experiments (**c**) were successfully reproduced at least three times. For gel source data, see Supplementary Fig. 1.

## Extended Data Fig. 2 Cryo-EM data processing workflow of telomerase with sstDNA (GTTGGG)<sub>5</sub> (telomerase T3D2) and the evaluation of the reconstruction.

**a**, Data processing workflow (detailed in Methods). Soft masks used in data processing are coloured in orange. **b**, Euler angle distributions of telomerase particles used for the 3.3 Å-resolution reconstruction. **c**, Local resolution evaluation of the 3.3 Å resolution cryo-EM map shown in surface views (left) and a slice view of the core region (right). **d**, Superposition of reconstructions P1, P2 and P3 that illustrates the rotation of TEN–TRAP. The three maps were low-pass-filtered to 6 Å and aligned on the TERT ring. p50 (red) and TEB bind to and move together with TEN–TRAP. **e**, Plot of the FSC as a function of the spatial frequency, with resolution of the final reconstruction indicated. **f**, FSC coefficients as a function of spatial frequency between model and cryo-EM density maps. Red curve, refined model versus half map 1 used for refinement; green curve, refined model versus half map 2 not used for refinement; black curve, refined model *versus* the combined final map. The generally similar appearances between the red and green curves suggests no substantial overfitting. **g**, Representative cryo-EM densities (grey and mesh) encasing the related atomic models (colour sticks and ribbons).

## Extended Data Fig. 3 Detailed interactions between TERT and p65 with TER.

**a**, Close-up view of motif 3N (amino acids 550–560). Motif 3A helix is bent towards motif 2, and motif 3N in between forms a finger-shaped architecture. **b**, Ribbon diagram of the TERT–TER ‘interlock’ with TERT domains coloured as indicated. **c**, Schematic of stem-loop 2, TBE and TBE<sub>L</sub> nucleotides and their interactions with RBD of TERT. Arrows indicate sites of polar interactions. Bold line represents the stacking interaction between Phe242 and C<sub>39</sub>. **d–f**, Structure of TER loop 4 and its interactions with RBD and CTE of TERT and xRRM of p65 (green). **g**, Rainbow-coloured ribbon diagram of La motif of p65 with secondary structural elements labelled. Positively charged and aromatic residues located on the interface between the La motif of p65 and TER are shown as spheres. **h**, Electrostatic surface representation of the La motif of p65 and its interactions with TER stem 1, pseudoknot and the 3' poly-U. The La motifs of p65 in **g** and **h** are in the same orientation. **i**, Schematic of pseudoknot with regions that interact with TERT and p65 indicated. **j**, Interactions between motif 3 and the template. End of motif 3B and start of motif 3C are in the minor groove of the duplex. **k**, The eight TER nucleotides that stack inside the catalytic cavity. Cryo-EM densities are shown as transparent meshes. Ideal A-form stacking of eight nucleotides (white) is shown for comparison. Backbone of the final three TER nucleotides in the stacking deviate from ideal A-form conformation.

## Extended Data Fig. 4 Interactions between TEN–TRAP and telomerase activity assays.

**a**, Ribbon representation of TERT with its domains coloured as indicated. Unmodelled regions of TERT are shown as dashed lines, including the linker between TEN and RBD (amino acids 180–215), flexible linkers within RBD (amino acids 252–280), and TRAP (amino acids 664–686). **b**, Hydrophobic interactions between the distal region of TRAP and the C-terminal helix of TEN domain, which is further stabilized by Gln168 via two hydrogen bonds. **c**, The extended β-sheet across TEN and TRAP. V791Y (Val791 in human corresponds to Val731 in *Tetrahymena*) mutation

in human TERT that disrupts telomere length maintenance and cell immortalization is located at the interface<sup>69</sup>. **d, e**, In vitro-reconstituted telomerase activity assays with TERT mutations on the TEN–TRAP interface. The top panels are SDS–PAGE gels showing the expression level of <sup>35</sup>S-Met incorporated TERT mutants. Quantification of activity and RAP for each mutant are shown in bar graphs below. **f, g**, Quantification of activity and RAP for gel shown in Figs. 1j, 3g. Data are mean ± s.d. from three independent experiments.

**Extended Data Fig. 5 Comparison between TERT from *Tetrahymena* and human, and the TERT-like protein from *Tribolium castaneum*.**

**a**, Sequence alignment of *Tetrahymena* TERT (TtTERT) human TERT (hTERT). Secondary structures and conserved motifs of *Tetrahymena* TERT are shown on top, with unmodelled regions shown as dashed lines. The alignments of the TEN, RBD, reverse transcriptase and CTE domains and TRAP motif were conducted separately with NIH COBALT<sup>70</sup> and then merged together. The alignment of CP2 and TFLY region was adjusted manually according to the previously reported alignment<sup>37</sup>. **b**, Structural comparison of the TERT-ring of *Tetrahymena* TERT (colour) and *Tribolium* TERT-like protein (grey) (PDB 3KYL). *Tribolium* TERT-like protein lacks TEN, TRAP and TER, and was crystallized with an artificial template–DNA duplex. **c, d**, Ribbon diagrams of template–DNA duplexes and surrounding structural elements of *Tetrahymena* TERT (**c**) and *Tribolium* TERT-like protein (**d**). The palm, fingers, primer grip, TH, TL, motif 3 and T are structurally conserved between *Tetrahymena* TERT and *Tribolium* TERT-like protein. The ‘bridge loop’ of *Tribolium* TERT-like protein is in a similar position to that in *Tetrahymena* TERT; however, the tip residues (Ser82 and Phe83) have no contact with the template–DNA duplex. CP2, which participates in template 5' boundary definition and template nucleotide guidance in *Tetrahymena* TERT, appears to be absent in *Tribolium* TERT-like protein.

**Extended Data Fig. 6 Details of p50 OB–TERT and Teb1C–sstDNA interactions.**

**a–d**, p50 OB–TERT interactions. **a**, Rainbow-coloured ribbon diagram of p50 OB with secondary structure elements labelled. **b**, Comparison of p50 OB (red) and human TPP1 OB (grey) (PDB 2I46) structures. **c**, TEN loop (amino acids 121–126) passes through a hydrophobic cleft of p50 OB. This loop is a disordered loop in the TEN-domain crystal structure<sup>71</sup>. **d**, Structure-based sequence alignment of p50 OB and human TPP1 OB. The secondary structure elements of p50 OB (red) and TPP1 OB (grey) are shown above and below the sequence alignment, respectively. Residues located at the interface between p50 OB and TERT are highlighted in yellow. The NOB and TEL patch residues in human TPP1 OB are indicated and coloured in yellow. The phosphorylation site Ser111 of TPP1 OB is coloured in green. Scaffold residues of L<sub>α2–β4</sub> shown in Fig. 2a (bottom) are coloured in blue. **e–h**, Teb1C–sstDNA interactions. **e**, Path of sstDNA from active site to Teb1C. Low-pass-filtered cryo-EM density of sstDNA (transparent surface) is superimposed with the unfiltered DNA density (green) to better show its flexible region from T<sub>20</sub> to G<sub>22</sub>. Cryo-EM densities corresponding to TERT domains, TER and Teb1C are coloured as in Fig. 1d. **f**, Sequence of the sstDNA used for the cryo-EM sample preparation with the template and Teb1C-interacting regions indicated. Nucleotides from G<sub>1</sub> to G<sub>16</sub> are invisible in the cryo-EM map. **g**, Interactions between sstDNA nucleotides and Teb1C as indicated in e. Intermolecular hydrogen bonds and stacking interactions are shown as dashed yellow lines and black arrows, respectively. **h**, Specific interactions between Teb1C residues Lys660 and Glu667 and sstDNA nucleotide G<sub>19</sub> are shown together with their cryo-EM densities. Hydrogen bonds and their lengths are indicated.

## Extended Data Fig. 7 Cryo-EM reconstructions of telomerase with different sstDNA bound.

**a**, List of sstDNA primers used for cryo-EM sample preparation and their sequences. DNA or LNA nucleotides that pair with the template are underlined. **b**, Resolution of reconstructions determined by gold-standard FSC at the 0.143 criterion. **c**, **d**, Cryo-EM data processing workflow of telomerase T4D4 and T5D5, and evaluations of the final reconstructions. Initial particle screening processes are analogous to those described in the

data processing workflow of telomerase T3D2 (Methods) and are omitted for brevity. Focused 3D classifications were performed to separate DNA-free and DNA-bound particles. Short duplexes were observed in both of T4D4 and T5D5 reconstructions. For telomerase T5D5, there is a subset of particles with a longer duplex that we attribute to the greater stability conferred on the duplex by LNA nucleotides at the thermodynamically most stable duplex ( $d\text{GGGGT}\cdot\text{rACCCC}$ ) formed in the previous step.

### Extended Data Fig. 8 Template–DNA duplexes in telomerase structures at different steps of telomeric DNA synthesis.

Top, sequences of sstDNA primers.  $T^L$  or  $G^L$  denotes LNA nucleotide. DNA or LNA nucleotides that pair with the template are underlined. Middle, ribbon diagrams of the duplex, template-adjacent nucleotides, bridge loop, TH and TL superimposed with cryo-EM densities (transparent surfaces). Bottom, schematics of the duplexes. The active site (red star), bridge loop residues (Arg413 and Phe414), and catalytic cavity (grey shade) in different structures are aligned to show the relative positions of the duplex. TER and DNA nucleotides are colour-coded as in Fig. 4.

### Extended Data Fig. 9 Structural details of template boundary determination (TBE, $\text{TBE}_L$ , $\text{TRE}_L$ and TRE) in telomerase T5D5.

**a**, Telomerase catalytic cavity in telomerase T5D5 with TER (grey) and DNA (green) shown as ribbon and TERT shown as surface (coloured). TBE,  $\text{TBE}_L$ , template,  $\text{TRE}_L$  and TRE nucleotides are highlighted as indicated. **b–e**, Detailed interactions between TERT and TER in regions as indicated in **a**. Intermolecular hydrogen bonds and stacking interactions are shown as dashed yellow lines and black arrows, respectively. The electrostatic surface of the TRAP–TH channel is shown in **d**. **f**, Schematic showing specific interactions between TERT and  $\text{TRE}_L$ –TRE as shown in **c**, **e**. Nucleotides from  $A_{54}$  to  $A_{58}$  are unmodelled and indicated as dashed orange lines. **g**, Predicted TRE and  $\text{TRE}_L$  conformation when the template is at the +1 position (template nucleotide  $C_{48}$  at the active site).  $\text{TRE}_L$

nucleotides C<sub>56</sub>, U<sub>57</sub> and A<sub>58</sub> would be fully stretched (about 5–6 Å phosphate-to-phosphate distance for each nucleotide) to span the distance from the neck of the TRAP–TH channel to the anchored TRE.

## **Extended Data Table 1 Cryo-EM data collection, refinement and validation statistics**

[Full size table](#)

## **Supplementary information**

### [\*\*Supplementary Figure 1\*\*](#)

This file contains Supplementary Figure 1.

### [\*\*Reporting Summary\*\*](#)

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

### [\*\*Supplementary Video 1\*\*](#)

Overall structure of Tetrahymena telomerase

### [\*\*Supplementary Video 2\*\*](#)

Architecture of the TERT–TER interlock.

## **Rights and permissions**

### [\*\*Reprints and Permissions\*\*](#)

## **About this article**



Check for  
updates

## Cite this article

He, Y., Wang, Y., Liu, B. *et al.* Structures of telomerase at several steps of telomere repeat synthesis. *Nature* **593**, 454–459 (2021).  
<https://doi.org/10.1038/s41586-021-03529-9>

### [Download citation](#)

- Received: 24 November 2020
- Accepted: 09 April 2021
- Published: 12 May 2021
- Issue Date: 20 May 2021
- DOI: <https://doi.org/10.1038/s41586-021-03529-9>

## Comments

By submitting a comment you agree to abide by our [Terms](#) and [Community Guidelines](#). If you find something abusive or that does not comply with our terms or guidelines please flag it as inappropriate.

[Access through your institution](#)

[Change institution](#)

[Buy or subscribe](#)

Advertisement

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

- Article
- [Published: 05 May 2021](#)

# A lysine–cysteine redox switch with an NOS bridge regulates enzyme function

- [Marie Wensien](#) ORCID: [orcid.org/0000-0003-4111-6918<sup>1,2 na1</sup>](https://orcid.org/0000-0003-4111-6918),
- [Fabian Rabe von Pappenheim](#) ORCID: [orcid.org/0000-0002-3925-8127<sup>1,2 na1</sup>](https://orcid.org/0000-0002-3925-8127),
- [Lisa-Marie Funk](#) ORCID: [orcid.org/0000-0003-2443-691X<sup>1,2</sup>](https://orcid.org/0000-0003-2443-691X),
- [Patrick Kloskowski](#)<sup>1,2</sup>,
- [Ute Curth](#)<sup>3</sup>,
- [Ulf Diederichsen](#)<sup>4</sup>,
- [Jon Uranga](#) ORCID: [orcid.org/0000-0002-1256-6656<sup>5</sup>](https://orcid.org/0000-0002-1256-6656),
- [Jin Ye](#)<sup>5</sup>,
- [Pan Fang](#)<sup>6</sup>,
- [Kuan-Ting Pan](#) ORCID: [orcid.org/0000-0001-6974-5324<sup>6</sup>](https://orcid.org/0000-0001-6974-5324),
- [Henning Urlaub](#) ORCID: [orcid.org/0000-0003-1837-5233<sup>6,7</sup>](https://orcid.org/0000-0003-1837-5233),
- [Ricardo A. Mata](#)<sup>5</sup>,
- [Viktor Sautner](#)<sup>1,2</sup> &
- [Kai Tittmann](#) ORCID: [orcid.org/0000-0001-7891-7108<sup>1,2</sup>](https://orcid.org/0000-0001-7891-7108)

[Nature](#) volume 593, pages 460–464 (2021) [Cite this article](#)

- 13k Accesses
- 329 Altmetric
- [Metrics details](#)

## Subjects

- [Enzyme mechanisms](#)
- [Proteins](#)
- [X-ray crystallography](#)

## Abstract

Disulfide bonds between cysteine residues are important post-translational modifications in proteins that have critical roles for protein structure and stability, as redox-active catalytic groups in enzymes or allosteric redox switches that govern protein function<sup>1,2,3,4</sup>. In addition to forming disulfide bridges, cysteine residues are susceptible to oxidation by reactive oxygen species, and are thus central not only to the scavenging of these but also to cellular signalling and communication in biological as well as pathological contexts<sup>5,6</sup>. Oxidized cysteine species are highly reactive and may form covalent conjugates with, for example, tyrosines in the active sites of some redox enzymes<sup>7,8</sup>. However, to our knowledge, regulatory switches with covalent crosslinks other than disulfides have not previously been demonstrated. Here we report the discovery of a covalent crosslink between a cysteine and a lysine residue with a NOS bridge that serves as an allosteric redox switch in the transaldolase enzyme of *Neisseria gonorrhoeae*, the pathogen that causes gonorrhoea. X-ray structure analysis of the protein in the oxidized and reduced state reveals a loaded-spring mechanism that involves a structural relaxation upon redox activation, which is propagated from the allosteric redox switch at the protein surface to the active site in the protein interior. This relaxation leads to a reconfiguration of key catalytic residues and elicits an increase in enzymatic activity of several orders of magnitude. The redox switch is highly conserved in related transaldolases from other members of the Neisseriaceae; for example, it is present in the transaldolase of *Neisseria meningitidis* (a pathogen that is the primary cause of meningitis and septicaemia in children). We surveyed the Protein Data Bank and found that the NOS bridge exists in diverse protein families across all domains of life (including *Homo sapiens*) and that it is often located at catalytic or regulatory hotspots. Our findings will inform strategies for the design of

proteins and peptides, as well as the development of new classes of drugs and antibodies that target the lysine–cysteine redox switch<sup>9,10</sup>.

## Access options

Subscribe to Journal

Get full journal access for 1 year

\$199.00

only \$3.90 per issue

[Subscribe](#)

All prices are NET prices.

VAT will be added later in the checkout.

Tax calculation will be finalised during checkout.

Rent or Buy article

Get time limited or full article access on ReadCube.

from \$8.99

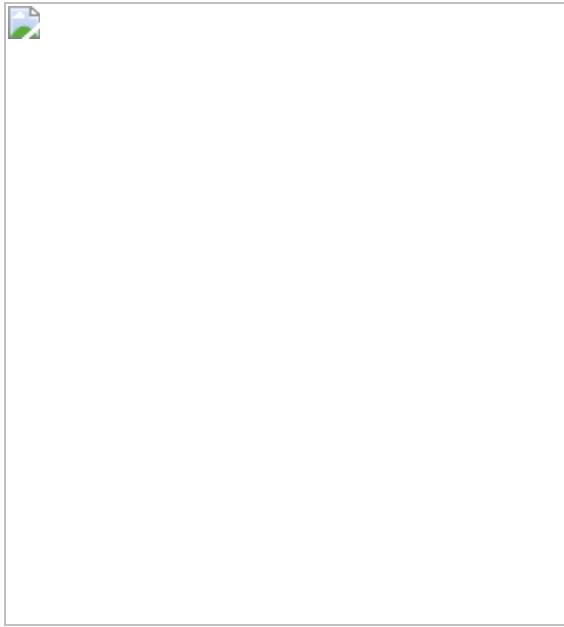
[Rent or Buy](#)

All prices are NET prices.

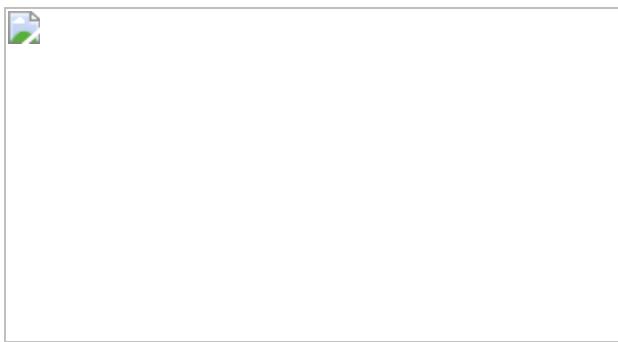
### Additional access options:

- [Log in](#)
- [Access through your institution](#)
- [Learn about institutional subscriptions](#)

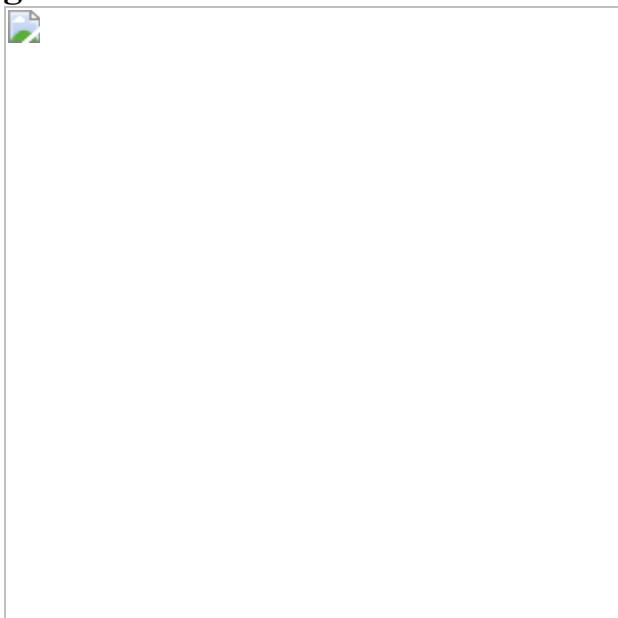
**Fig. 1: Functional and structural analysis of NgTAL in the oxidized and reduced state.**



**Fig. 2: Structure of the allosteric redox switch in NgTAL in the oxidized and reduced state.**



**Fig. 3: Structural basis of redox activation of NgTAL.**



## Data availability

The refined structural protein models and corresponding structure–factor amplitudes have been deposited under PDB accession codes [6XZ4](#) (*NgTAL oxidized citrate 1*), [6ZWJ](#) (*NgTAL oxidized citrate 2*), [6ZWH](#) (*NgTAL oxidized acetate*), [6ZWF](#) (*NgTAL reduced citrate*), [7B0L](#) (*NgTAL oxidized low-dose*), [7BBX](#) (*NgTAL Lys8Ala variant*) and [7BBW](#) (*NgTAL Cys38Ser variant*). The structures cited in this publication (1M3Q, 3CLM, 6T3X and 5Y72) are available under their respective PDB accession codes. The data for our protein database are currently private, and can only be accessed with a single reviewer account that has been created. The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier PXD020302 (reviewer account details: reviewer52532@ebi.ac.uk (username), AK3E73R2 (password)). All other data are available from the corresponding author on request.

## References

1. 1.

Bardwell, J. C., McGovern, K. & Beckwith, J. Identification of a protein required for disulfide bond formation in vivo. *Cell* **67**, 581–589 (1991).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

2. 2.

Wedemeyer, W. J., Welker, E., Narayan, M. & Scheraga, H. A. Disulfide bonds and protein folding. *Biochemistry* **39**, 4207–4216 (2000).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

3. 3.

Matsumura, M. & Matthews, B. W. Control of enzyme activity by an engineered disulfide bond. *Science* **243**, 792–794 (1989).

[ADS](#) [CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

4. 4.

Hogg, P. J. Disulfide bonds as switches for protein function. *Trends Biochem. Sci.* **28**, 210–214 (2003).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

5. 5.

Paulsen, C. E. & Carroll, K. S. Cysteine-mediated redox signaling: chemistry, biology, and tools for discovery. *Chem. Rev.* **113**, 4633–4679 (2013).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

6. 6.

Depuydt, M. et al. A periplasmic reducing system protects single cysteine residues from oxidation. *Science* **326**, 1109–1111 (2009).

[ADS](#) [CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

7. 7.

Ito, N. et al. Novel thioether bond revealed by a 1.7 Å crystal structure of galactose oxidase. *Nature* **350**, 87–90 (1991).

[ADS](#) [CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

8. 8.

Firbank, S. J. et al. Crystal structure of the precursor of galactose oxidase: an unusual self-processing enzyme. *Proc. Natl Acad. Sci. USA* **98**, 12932–12937 (2001).

[ADS](#) [CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

9. 9.

Dror, R. O. et al. Structural basis for modulation of a G-protein-coupled receptor by allosteric drugs. *Nature* **503**, 295–299 (2013).

[ADS](#) [CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

10. 10.

Nussinov, R. & Tsai, C. J. The design of covalent allosteric drugs. *Annu. Rev. Pharmacol. Toxicol.* **55**, 249–267 (2015).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

11. 11.

Kang, H. J., Coulibaly, F., Clow, F., Proft, T. & Baker, E. N. Stabilizing isopeptide bonds revealed in Gram-positive bacterial pilus structure. *Science* **318**, 1625–1628 (2007).

[ADS](#) [CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

12. 12.

Uhlén, U. & Eklund, H. Structure of ribonucleotide reductase protein R1. *Nature* **370**, 533–539 (1994).

[ADS](#) [CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

13. 13.

Lehwess-Litzmann, A. et al. Twisted Schiff base intermediates and substrate locale revise transaldolase mechanism. *Nat. Chem. Biol.* **7**, 678–684 (2011).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

14. 14.

Tittmann, K. Sweet siblings with different faces: the mechanisms of FBP and F6P aldolase, transaldolase, transketolase and phosphoketolase revisited in light of recent structural data. *Bioorg. Chem.* **57**, 263–280 (2014).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

15. 15.

Wi, T. et al. Antimicrobial resistance in *Neisseria gonorrhoeae*: global surveillance and a call for international collaborative action. *PLoS Med.* **14**, e1002344 (2017).

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

16. 16.

Chiu, J. & Hogg, P. J. Allosteric disulfides: sophisticated molecular structures enabling flexible protein regulation. *J. Biol. Chem.* **294**, 2949–2960 (2019).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

17. 17.

Neumann, P. & Tittmann, K. Marvels of enzyme catalysis at true atomic resolution: distortions, bond elongations, hidden flips, protonation states and atom identities. *Curr. Opin. Struct. Biol.* **29**, 122–133 (2014).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

18. 18.

Ruszkowski, M. & Dauter, Z. On methylene-bridged cysteine and lysine residues in proteins. *Protein Sci.* **25**, 1734–1736 (2016).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

19. 19.

Wang, J. Crystallographic identification of spontaneous oxidation intermediates and products of protein sulphydryl groups. *Protein Sci.* **28**, 472–477 (2019).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

20. 20.

Kovacic, P. & Bennett, R. P. Aromatic amination with hydroxylamine-*O*-sulfonic acid. *J. Am. Chem. Soc.* **83**, 221–224 (1961).

[CAS](#) [Google Scholar](#)

21. 21.

Akter, S. et al. Chemical proteomics reveals new targets of cysteine sulfenic acid reductase. *Nat. Chem. Biol.* **14**, 995–1004 (2018).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

22. 22.

Kirby, A. J. et al. Hydroxylamine as an oxygen nucleophile. Structure and reactivity of ammonia oxide. *J. Am. Chem. Soc.* **128**, 12374–12375 (2006).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

23. 23.

Lüdtke, S. et al. Sub-Ångström-resolution crystallography reveals physical distortions that enhance reactivity of a covalent enzymatic intermediate. *Nat. Chem.* **5**, 762–767 (2013).

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

24. 24.

Mesecar, A. D., Stoddard, B. L. & Koshland, D. E. Jr. Orbital steering in the catalytic power of enzymes: small structural changes with large catalytic consequences. *Science* **277**, 202–206 (1997).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

25. 25.

Schörken, U. et al. Identification of catalytically important residues in the active site of *Escherichia coli* transaldolase. *Eur. J. Biochem.* **268**, 2408–2415 (2001).

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

26. 26.

Dai, S. et al. Low-barrier hydrogen bonds in enzyme cooperativity. *Nature* **573**, 609–613 (2019).

[ADS](#) [CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

27. 27.

Rabe von Pappenheim, F. et al. Structural basis for antibiotic action of the B<sub>1</sub> antivitamin 2'-methoxy-thiamine. *Nat. Chem. Biol.* **16**, 1237–1245 (2020).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

28. 28.

Seib, K. L. et al. Defenses against oxidative stress in *Neisseria gonorrhoeae*: a system tailored for a challenging environment. *Microbiol. Mol. Biol. Rev.* **70**, 344–361 (2006).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

29. 29.

Seib, K. L., Tseng, H. J., McEwan, A. G., Apicella, M. A. & Jennings, M. P. Defenses against oxidative stress in *Neisseria gonorrhoeae* and *Neisseria meningitidis*: distinctive systems for different lifestyles. *J. Infect. Dis.* **190**, 136–147 (2004).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

30. 30.

Bruner, S. D., Norman, D. P. & Verdine, G. L. Structural basis for recognition and repair of the endogenous mutagen 8-oxoguanine in DNA. *Nature* **403**, 859–866 (2000).

[ADS](#) [CAS](#) [PubMed](#) [Google Scholar](#)

31. 31.

Chevillard, S. et al. Mutations in *OGG1*, a gene involved in the repair of oxidative DNA damage, are found in human lung and kidney tumours. *Oncogene* **16**, 3083–3086 (1998).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

32. 32.

Huang, P. S., Boyken, S. E. & Baker, D. The coming of age of de novo protein design. *Nature* **537**, 320–327 (2016).

[ADS](#) [CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

33. 33.

Singh, J., Petter, R. C., Baillie, T. A. & Whitty, A. The resurgence of covalent drugs. *Nat. Rev. Drug Discov.* **10**, 307–317 (2011).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

34. 34.

Scott, A. M., Wolchok, J. D. & Old, L. J. Antibody therapy of cancer. *Nat. Rev. Cancer* **12**, 278–287 (2012).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

35. 35.

Muyldermans, S. Nanobodies: natural single-domain antibodies. *Annu. Rev. Biochem.* **82**, 775–797 (2013).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

36. 36.

Senior, A. W. et al. Improved protein structure prediction using potentials from deep learning. *Nature* **577**, 706–710 (2020).

[ADS](#) [CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

37. 37.

Callaway, E. ‘It will change everything’: DeepMind’s AI makes gigantic leap in solving protein structures. *Nature* **588**, 203–204 (2020).

[ADS](#) [CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

38. 38.

Gill, S. C. & von Hippel, P. H. Calculation of protein extinction coefficients from amino acid sequence data. *Anal. Biochem.* **182**, 319–326 (1989).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

39. 39.

Inoue, H., Nojima, H. & Okayama, H. High efficiency transformation of *Escherichia coli* with plasmids. *Gene* **96**, 23–28 (1990).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

40. 40.

Bertani, G. Studies on lysogenesis. I. The mode of phage liberation by lysogenic *Escherichia coli*. *J. Bacteriol.* **62**, 293–300 (1951).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

41. 41.

Tsolas, O. & Joris, L. Transaldolase. *Methods Enzymol.* **42**, 290–297 (1975).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

42. 42.

Sautner, V., Friedrich, M. M., Lehwess-Litzmann, A. & Tittmann, K. Converting transaldolase into aldolase through swapping of the multifunctional acid–base catalyst: common and divergent catalytic principles in F6P aldolase and transaldolase. *Biochemistry* **54**, 4475–4486 (2015).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

43. 43.

Schuck, P. Size-distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and lamm equation modeling. *Biophys. J.* **78**, 1606–1619 (2000).

[ADS](#) [CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

44. 44.

Laue, M. T., Shah, B. D., Rigideway, T. M. & Pelletier, S. L. in *Analytical Ultracentrifugation in Biochemistry and Polymer Science* (eds Harding S. et al.) 90–125 (Royal Society of Chemistry, 1992).

45. 45.

Brautigam, C. A. Calculations and publication-quality illustrations for analytical ultracentrifugation data. *Methods Enzymol.* **562**, 109–133 (2015).

[CAS](#) [Google Scholar](#)

46. 46.

Kabsch, W. Xds. *Acta Crystallogr. D* **66**, 125–132 (2010).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

47. 47.

Adams, P. D. et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D* **66**, 213–221 (2010).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

48. 48.

Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D* **66**, 486–501 (2010).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

49. 49.

Williams, C. J. et al. MolProbity: More and better reference data for improved all-atom structure validation. *Protein Sci.* **27**, 293–315 (2018).

[CAS](#) [Google Scholar](#)

50. 50.

Schrödinger. The PyMOL Molecular Graphics System, version 1.8 (2015).

51. 51.

Liebschner, D. et al. Polder maps: improving OMIT maps by excluding bulk solvent. *Acta Crystallogr. D* **73**, 148–157 (2017).

[CAS](#) [Google Scholar](#)

52. 52.

Frisch, M. J. et al. Gaussian 16, revision C.01 (Gaussian, 2016).

53. 53.

Becke, A. D. Density-functional thermochemistry. III. The role of exact exchange. *J. Chem. Phys.* **98**, 5648–5652 (1993).

[ADS](#) [CAS](#) [Google Scholar](#)

54. 54.

Grimme, S., Ehrlich, S. & Goerigk, L. Effect of the damping function in dispersion corrected density functional theory. *J. Comput. Chem.* **32**, 1456–1465 (2011).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

55. 55.

Rappoport, D. & Furche, F. Property-optimized Gaussian basis sets for molecular response calculations. *J. Chem. Phys.* **133**, 134105 (2010).

[ADS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

56. 56.

Wong, C. P. et al. Two distinct substrate binding modes for the normal and reverse prenylation of hapalindoles by the prenyltransferase AmbP3. *Angew. Chem. Int. Ed.* **57**, 560–563 (2018).

[CAS](#) [Google Scholar](#)

57. 57.

Muller, Y. A. et al. High-resolution crystal structures of two prototypical  $\beta$ - and  $\gamma$ -herpesviral nuclear egress complexes unravel the determinants of subfamily specificity. *J. Biol. Chem.* **295**, 3189–3201 (2020).

[CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

[Download references](#)

## Acknowledgements

This study was supported by the Max-Planck Society and the DFG-funded Göttingen Graduate Center for Neurosciences, Biophysics and Molecular Biosciences GGNB. We acknowledge access to beamline P14 at DESY/EMBL (Hamburg), and thank G. Bourenkov and T. Schneider for local support. We thank G. Bourenkov, A. Pearson, H. Stark, R. Kluger and M. Alcarazo for discussion; and A. Linden for help with the mass spectrometry experiments.

## Author information

### Author notes

1. These authors contributed equally: Marie Wensien, Fabian Rabe von Pappenheim

## Affiliations

1. Department of Molecular Enzymology, Göttingen Center of Molecular Biosciences, Georg August University Göttingen, Göttingen, Germany

Marie Wensien, Fabian Rabe von Pappenheim, Lisa-Marie Funk, Patrick Kloskowski, Viktor Sautner & Kai Tittmann

2. Department of Structural Dynamics, Max Planck Institute for Biophysical Chemistry, Göttingen, Germany

Marie Wensien, Fabian Rabe von Pappenheim, Lisa-Marie Funk, Patrick Kloskowski, Viktor Sautner & Kai Tittmann

3. Institute for Biophysical Chemistry, Hannover Medical School, Hannover, Germany

Ute Curth

4. Institute for Organic and Biomolecular Chemistry, Georg August University Göttingen, Göttingen, Germany

Ulf Diederichsen

5. Institute of Physical Chemistry, Georg August University Göttingen, Göttingen, Germany

Jon Uranga, Jin Ye & Ricardo A. Mata

6. Bioanalytical Mass Spectrometry research group, Max Planck Institute for Biophysical Chemistry, Göttingen, Germany

Pan Fang, Kuan-Ting Pan & Henning Urlaub

7. University Medical Center Göttingen, Bioanalytics, Institute for Clinical Chemistry, Göttingen, Germany

Henning Urlaub

Authors

1. Marie Wensien

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

2. Fabian Rabe von Pappenheim

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

3. Lissa-Marie Funk

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

4. Patrick Kloskowski

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

5. Ute Curth

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

6. Ulf Diederichsen

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

7. Jon Uranga

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

8. Jin Ye

[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

9. Pan Fang  
[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

10. Kuan-Ting Pan  
[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

11. Henning Urlaub  
[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

12. Ricardo A. Mata  
[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

13. Viktor Sautner  
[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

14. Kai Tittmann  
[View author publications](#)

You can also search for this author in [PubMed](#) [Google Scholar](#)

## Contributions

K.T. designed and coordinated research; M.W. expressed, purified, biochemically characterized and crystallized *NgTAL* proteins with initial support from P.K.; L.-M.F. expressed, purified and biochemically characterized *NmTAL*; F.R.v.P. and V.S. conducted crystallographic data collection, processing, model building and refinement with support from K.T.; F.R.v.P., V.S. and K.T. interpreted crystallographic data; F.R.v.P. did

phylogenetic analysis; U.C. conducted and analysed analytical ultracentrifugation experiments; P.F. and K.-T.P. performed and analysed mass spectrometry experiments under the supervision of H.U.; R.A.M. supervised and planned theoretical investigations; J.U. performed electronic structure calculations including potentials,  $pK_a$  value and structures of reaction intermediates; J.Y. performed electronic structure calculations of reaction intermediates and pathways; U.D., J.U., R.A.M. and K.T. discussed reaction mechanisms; K.T. wrote the manuscript with input from all authors.

## Corresponding author

Correspondence to [Kai Tittmann](#).

## Ethics declarations

## Competing interests

M.W., F.R.v.P., L.-M.F., U.D., J.U., R.A.M., V.S. and K.T. have filed a European patent application (application number EP21164101.4) for regulating protein activities by targeting the NOS redox switches.

## Additional information

**Peer review information** *Nature* thanks Deborah Fass, Martin Högbom and Judith P. Klinman for their contribution to the peer review of this work. Peer reviewer reports are available.

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

## Extended data figures and tables

[Extended Data Fig. 1 Redox chemistry of cysteine residues in proteins and electron density maps, showing the presence of the](#)

## Lysine–cysteine NOS bridge in independent crystal structures of NgTAL.

**a**, Redox reactions of cysteine with key species involved, including cysteine oxidation by reactive oxygen species (ROS) and reduction by thiol antioxidants (RSH) or enzymes (Srx). **b**, Mechanism of allosteric redox switch containing an intramolecular allosteric disulfide bridge. Reduction of the disulfide to the corresponding dithiol results in a structural reorganization of the protein that is propagated to the active-site altering activity. **c**, NgTAL crystal 1. Top left,  $2mF_o - DF_c$  electron density map after final refinement contoured at  $3\sigma$ . No additional positive or negative electron density was observed at  $\pm 3\sigma$ . Top right,  $mF_o - DF_c$  omit electron density map contoured at  $5\sigma$ . Bottom, alternative refinement with a methylene bridge yields unexplained positive difference electron density at the bridging atom. The  $2mF_o - DF_c$  map is shown in grey at  $3\sigma$ , positive (green) and negative (red) electron density maps are shown at  $3\sigma$  and  $-3\sigma$ , respectively. **d**, NgTAL crystal 2. Left,  $2mF_o - DF_c$  electron density map after final refinement, contoured at  $1.5\sigma$ . Right,  $mF_o - DF_c$  omit electron density map contoured at  $5\sigma$ . **e**, NgTAL crystal 3. Left,  $2mF_o - DF_c$  electron density map after final refinement, contoured at  $1.5\sigma$ . Right,  $mF_o - DF_c$  omit electron density map contoured at  $5\sigma$ . **f**, NgTAL crystal 4. Low-dose dataset of an independent crystal measured in-house at a rotating anode. The NOS bridge is also present in the corresponding NgTAL structure (left,  $2mF_o - DF_c$  electron density map contoured at  $1\sigma$ ; right,  $mF_o - DF_c$  omit electron density map contoured at  $3\sigma$ ), thus ruling out that formation of the covalent linkage seen in crystals 1–3 results from radiation damage at the high-energy synchrotron beamline. **g**, Structure of the Lys8–Cys38 redox switch in the oxidized (left) and reduced (right) state, showing electron density for neighbouring waters (W1 and W2) and a presumed dioxygen molecule ( $O_2$ ) that is exclusively observed in the reduced state. The corresponding  $2mF_o - DF_c$  electron density map is shown in blue at a contour level of  $1.5\sigma$ .

## Extended Data Fig. 2 Thermal unfolding, analytical ultracentrifugation and X-ray structures of NgTAL Lys8Ala

## and Cys38Ser variants.

**a**, Thermal unfolding of NgTAL wild type and the Lys8Ala and Cys38Ser variants under oxidizing and reducing conditions, as monitored by far-UV CD spectroscopy at 222 nm. Different unfolding temperatures are seen for the oxidized and reduced states in case of the wild-type enzyme and the Lys8Ala variant, whereas the Cys38Ser variant does not exhibit this feature. This suggests an oxidation of Cys38 in the Lys8Ala variant, despite the absence of Lys8. **b**, Analytical ultracentrifugation analyses of NgTAL wild-type and variants in the oxidized and reduced state shows the predominant formation of the monomeric form in all of the cases we tested. Under oxidizing conditions and high protein concentrations, a small fraction of higher oligomers is observed (presumably resulting from incorrectly linked monomers). **c**, X-ray crystallographic structure of the NgTAL Lys8Ala variant, showing the allosteric redox switch site with residues Ala8 (mutation site), Cys38, Glu93 and Thr101. For residues Ala8 and Cys38, the corresponding  $2mF_o - DF_c$  electron density maps are shown in blue at a contour level of  $2\sigma$ . Inset, peaks in the  $mF_o - DF_c$  difference electron density map (in green, contour level  $3\sigma$ ) around the sulfur atom of residue Cys38 suggest that this atom is oxidized. Owing to the structural flexibility of Cys38, the discrete oxidation state (mono-oxidized and/or dioxidized) cannot be unambiguously assigned. Notwithstanding this ambiguity, this observation supports our proposed mechanism of an initial cysteine oxidation as part of the formation of the NOS bridge. **d**, X-ray crystallographic structure of the NgTAL Cys38Ser variant, showing the allosteric redox switch site with residues Lys8, Ser38 (mutation site), Glu93 and Thr101. For residues Lys8 and Ser38, the corresponding  $2mF_o - DF_c$  electron density maps are shown in blue at a contour level of  $1.5\sigma$ . Lys8 is chemically unmodified, thus ruling out that the covalent linkage between Lys8 and Cys38 seen in the wild-type enzyme results from the addition of CO<sub>2</sub> or formaldehyde potentially establishing an NCS linkage<sup>18</sup>.

## Extended Data Fig. 3 Putative reaction mechanisms of lysine-cysteine NOS bridge formation and associated computational calculations.

**a**, Initial reaction of the cysteine thiolate with dioxygen and subsequent attack of the lysine amine onto the  $\alpha$ -oxygen atom of the thio-(hydro)peroxy intermediate concomitantly with proton transfers and water release. **b**, Oxidation of both cysteine and lysine in either concerted fashion (top path) or independently (bottom path), followed by nucleophilic attack of the oxidized lysine as a O-nucleophile onto the cysteine sulfenic acid with concomitant water release. **c**, Initial attack of the lysine amine onto the sulfur atom of sulfenic acid or sulfinic acid to afford a sulfinamide species followed by [1,2] rearrangement driven by orbital steering. **d**, Depiction of selected reaction intermediates and relative free Gibbs energies ( $T = 298.15\text{ K}$ ) computed at the B3LYP-D3(BJ)/def2-TZVPD//B3LYP-D3(BJ)/def2-SVPD level of theory (as described in ‘Computational details’ in Methods). The reference is given by the sulfinic acid state (the thermodynamically most stable intermediate in our investigations). Top, starting thio-(hydro)peroxy species, which bears a barrier of 12.2 kcal mol $^{-1}$  for the heterolytic cleavage of the O–O bond and concurrent oxidation of the lysine residue. We compare two different pathways (using the same nomenclature as in **a–c**) through the sulfenic and sulfinic acids. The mechanism in **a** is not directly depicted, as we have not observed any concerted O–O cleavage with amino nucleophilic attack and NOS bridge formation. Only the mechanisms in **b**, **c** are depicted. A path that leads through the sulfinic acid or sulfinamide species is not viable, because the two species are far too stable. Instead, we suggest that the formal oxidation state of sulfur [0] is kept, with concurrent oxidation of the Lys8 (pathway from **b**). This would represent a reversible mechanism. For each intermediate, three different protonation states were investigated with a total cluster charge of –1, 0 and 1. Only the most stable species are shown.

#### Extended Data Fig. 4 Phylogenetic analyses and sequence conservation of the lysine–cysteine redox switch in the TAL protein family using NgTAL as reference.

Two related consensus motifs were identified that contain the lysine and cysteine residues of the redox switch, the active site serine and asparagine residues required for catalytic activation of the Schiff-base-forming lysine, and the linker region that connects the redox switch with the active site. The identified motifs are highly conserved in Betaproteobacteria (in particular,

Neisseriales (motif 1), and in Cyanophycea (motif 2)). *NmTAL* is highly similar (95% identity) to *NgTAL*, and also contains motif 1. Structural and functional analyses of *NmTAL* could confirm the existence of a redox switch (Extended Data Fig. 5).

**Extended Data Fig. 5 Structural and functional analyses of *NmTAL* indicate the presence of a redox switch.**

**a**, Steady-state kinetic analysis of enzymatic activity of *NmTAL* in the oxidized (black) and reduced (red) state. There is a multi-fold increase of  $k_{\text{cat}}$  and concomitant decrease of substrate  $K_M$  upon reduction. The catalytic constant of the oxidized form represents an upper limit, as oxidized and reduced species cannot be quantitatively separated by chromatographic methods as in the case of *NgTAL*. All measurements were carried out in triplicate and are shown as mean  $\pm$  s.d. Kinetic and thermodynamic constants are provided in Extended Data Table 1. Experiments were repeated twice with similar results. **b**, Far-UV CD spectra of *NmTAL* in the oxidized (black) and reduced (red) state, showing both the natively folded states (solid lines) and states after thermal unfolding (dashed lines). The reduced enzyme completely unfolds and does not contain residual secondary structure. By contrast, the oxidized enzyme exhibits only partial unfolding of mostly helical elements and contains thermally stable  $\beta$ -sheet structures. **c**, Thermal unfolding of *NmTAL* in the oxidized (black) and reduced (red) state, monitored by far-UV CD spectroscopy at 222 nm. Although the oxidized enzyme displays a monophasic unfolding with a melting temperature of 50.6 °C corresponding to unfolding of the  $\alpha$ -helices (as shown in b), the reduced enzyme shows a biphasic unfolding with melting temperatures of 51.9 °C (first transition) and 71.9 °C (second transition). This observation suggests that the putative NOS bridge in the oxidized state of *NmTAL* specifically stabilizes the interior  $\beta$ -sheet structure of the TIM barrel, in agreement with the position of switch residue Cys38.

**Extended Data Fig. 6 Structure of the human DNA repair enzyme 8-oxoguanine glycosylase 1 with a putative NOS bridge at the active site.**

Structure is from PDB 1M3Q, and is at 1.90 Å resolution. **a**, Overall structure of human 8-oxoguanine glucosylase 1 in complex with DNA and product analogue 8-aminoguanine. **b**, Close-up view of the active site, showing the product and residues Cys253 and Lys249. Residues Cys253 and Lys249 are superposed with the calculated  $2mF_o - DF_c$  electron density map (in blue, contour level  $1.5\sigma$ ). A strong, unexplained positive peak in the  $mF_o - DF_c$  difference electron density map (in green, contour level  $3\sigma$ ) is observed between the cysteine sulfur atom and the lysine nitrogen atom. The S–N interatomic distance (2.80 Å in the previously deposited structure) is too short for a hydrogen-bond interaction and too long for a direct S–N linkage. **c**, Refinement of the previously deposited structure with a geometrically parametrized NOS bridge linking Cys253 and Lys249 resulted in a structural model with no remaining unexplained electron density ( $2mF_o - DF_c$  electron density map in blue, contour level  $1.5\sigma$ ;  $mF_o - DF_c$  difference electron density map in green and red, contour level  $\pm 3\sigma$ ), suggesting the presence of an NOS bridge.

**Extended Data Fig. 7 Representative examples of protein structures deposited in the PDB that probably contain an NOS bridge akin to that of NgTAL.**

**a**, Prenyltransferase AmbP3. **b**, Human cytomegalovirus pUL50–pUL53 complex. For both examples, the overall structure is shown in cartoon representation in the top panel, highlighting the Lys–Cys linkage and providing the PDB code<sup>56,57</sup>. In the corresponding middle panels, the structure of the lysine–cysteine pair as deposited in the PDB is shown enlarged, including the calculated  $2mF_o - DF_c$  (in blue, contour level  $1\sigma$ ) and  $mF_o - DF_c$  difference (in green, contour level  $3\sigma$ ) electron density maps. There is a pronounced positive difference peak in the electron density maps in between the lysine nitrogen atoms and the cysteine sulfur atoms, indicating the presence of a covalent bridge. In the bottom panels, the refined structural models that include the covalent lysine–cysteine NOS bridges are shown with the corresponding  $2mF - DF_c$  electron density maps. The  $mF_o - DF_c$  difference electron density maps are shown in green and are contoured at  $3\sigma$ . The calculated occupancies of the NOS bridges

amount to 62% (**a**) and 76% (**b**). The NOS bridge is prominently located at either the substrate binding site or the protein–protein binding interface.

### **Extended Data Fig. 8 Superposition of NgTAL in the oxidized and reduced state.**

Both of these structures are from this Article. **a**, Structure of the active site of NgTAL in the oxidized state, showing catalytic residues Lys138, Asp17 and the ligand citrate. Citrate is partially disordered. The corresponding  $2mF_o - DF_c$  (blue) and  $mF_o - DF_c$  omit (green) electron density maps are shown at contour levels of  $1\sigma$  and  $5\sigma$ , respectively. **b**, Structure of the active site of NgTAL in the reduced state, showing catalytic residues Lys138, Asp17 and the ligand citrate. Citrate is structurally well-defined in this state. A covalent conjugate between Lys138 and a two-carbon fragment formed during crystallization. The corresponding  $2mF_o - DF_c$  (blue) and  $mF_o - DF_c$  omit (green) electron density maps are shown at contour levels of  $1\sigma$  and  $5\sigma$ , respectively. **c**, Superposition of the oxidized (yellow) and reduced (grey) NgTAL, showing selected active site residues and the ligand citrate. The structural change of ligand citrate is accompanied by a redistribution of the two conformers of Arg204 (oxidized 70:30% occupancy, reduced 30:70% occupancy). There is also a subtle repositioning of the active-site residues (for example, Asp17 and water molecules).

### **Extended Data Table 1 Steady-state kinetic analysis of wild-type NgTAL and its variants, as well as of NmTAL wild type, under oxidizing and reducing conditions**

[Full size table](#)

### **Extended Data Table 2 X-ray crystallographic data collection and refinement statistics**

[Full size table](#)

## **Supplementary information**

### **Supplementary Information**

This file contains Supplementary Results: Mass spectrometric analyses of NgTAL in the reduced and oxidized state; Supplementary Fig. 1: Chromatographic analysis of NgTAL in the reduced and oxidized state; and Supplementary Fig. 2: Mass spectrometric analysis of post-translational modifications of NgTAL under oxidizing and reducing conditions.

## [Reporting Summary](#)

## [Peer Review File](#)

## Rights and permissions

### [Reprints and Permissions](#)

## About this article



Check for  
updates

## Cite this article

Wensien, M., von Pappenheim, F.R., Funk, LM. *et al.* A lysine–cysteine redox switch with an NOS bridge regulates enzyme function. *Nature* **593**, 460–464 (2021). <https://doi.org/10.1038/s41586-021-03513-3>

### [Download citation](#)

- Received: 16 September 2020
- Accepted: 01 April 2021
- Published: 05 May 2021
- Issue Date: 20 May 2021

- DOI: <https://doi.org/10.1038/s41586-021-03513-3>

## Comments

By submitting a comment you agree to abide by our [Terms](#) and [Community Guidelines](#). If you find something abusive or that does not comply with our terms or guidelines please flag it as inappropriate.

[Access through your institution](#)

[Change institution](#)

[Buy or subscribe](#)

## Associated Content

### [Previously unknown type of protein crosslink discovered](#)

- Deborah Fass
- Sergey N. Semenov

Nature News & Views 05 May 2021

Advertisement

---

This article was downloaded by **calibre** from <https://www.nature.com/articles/s41586-021-03513-3>