

# nature



## THE SCIENCE OF INEQUALITY

How quantifying disparities can help to make societies fairer

**Net gains**  
Small-scale fisheries offer a sustainable path to food production

**Quickswitch**  
High-speed bipolar transistor formed from organic semiconductor

**Active diet**  
Exercise-induced molecule suppresses appetite in mice

# Nature.2022.06.25

[Sat, 25 Jun 2022]

- [The science of inequality](#)
- [This Week](#)
- [News in Focus](#)
- [Opinion](#)
- [Research](#)
- [Amendments & Corrections](#)
- [Nature Outlook](#)

# The science of inequality

- **[Equity must be baked into randomized controlled trials](#)** [ 22 Jun 2022]
  - Experiments are transforming development research – but scientists must ensure the benefits reach those most in need.
- **[I have funds to buy reagents, but not remedies](#)** [ 22 Jun 2022]
  - Ignoring the challenges of research in low-income countries only perpetuates inequity.
- **[How COVID has deepened inequality — in six stark graphics](#)** [ 22 Jun 2022]
  - Troubling data show how the pandemic has exacted an unequal toll, pushing tens of millions into poverty and having the greatest effects on already-disadvantaged groups.
- **[These experiments could lift millions out of dire poverty](#)** [ 22 Jun 2022]
  - Randomized trials are changing the way governments and aid organizations study — and deliver — measures to reduce inequality and poverty.
- **[Tackling inequality takes social reform](#)** [ 22 Jun 2022]
  - In separate books, leading economists explore the wide-ranging changes needed to produce a more just society.
- **[Not all inequalities are alike](#)** [ 22 Jun 2022]
  - Better data and new statistical techniques could enable researchers to measure the form of inequality that seems most harmful to society — inequality of opportunity.
- **[The rise of inequality research: can spanning disciplines help tackle injustice?](#)** [ 22 Jun 2022]
  - Wide-ranging expertise and direct involvement of those affected will help to make inequality research more meaningful.
- **[Shark researcher attacks lack of diversity in marine science](#)** [ 22 Jun 2022]
  - Jasmin Graham has a passion for predators and promoting inclusive science.

- EDITORIAL
- 22 June 2022

# Equity must be baked into randomized controlled trials

Experiments are transforming development research – but scientists must ensure the benefits reach those most in need.



Alexandra township in Johannesburg is next to Sandton, one of South Africa's wealthiest neighbourhoods. Credit: Chip Somodevilla/Getty

More than two decades ago, researchers began using rigorous experiments to test policies designed to improve the lives of some of the world's poorest

citizens. The movement grew, and randomized controlled trials (RCTs), once reserved for drugs and other health interventions, are now part of the mainstream in the research community that studies global development.

The knowledge resulting from RCTs of all kinds has helped people across the world. It has led to better drugs and new health interventions, as well as educational opportunities, improved agricultural technologies and effective programmes to distribute desperately needed cash. But these benefits have not necessarily flowed equitably, or helped those who need it the most.

The pernicious and pervasive effects of inequality in all its forms are all too clear, as *Nature* reports this week. A [special collection of articles](#) includes content that looks at the ways in which COVID-19 has deepened inequality; factors that contribute to gender inequality in science; and an analysis of interventions designed to alleviate poverty.

Researchers are working to understand and measure the root causes, as well as the symptoms, of social inequities. At the same time, like so many others, they are also confronting their own role in a complicated socio-economic system in which benefits such as wealth and opportunity too often flow to those who already experience such advantages.

One of the first things that researchers can do is to ensure that their work is not contributing to outstanding social inequities. Sadly, all too often, the benefits of public-health research are distributed inequitably. Health interventions ranging from COVID-19 vaccines or drugs to new technologies can improve overall public health, but if they reach people in rich countries, or people who already have access to high-quality health care, before those who have fewer advantages, they [will also contribute to inequality](#).

Furthermore, even when the interventions being tested aim to relieve poverty, they don't necessarily have that effect. Many studies show that the poorest and most disadvantaged people often benefit the most from RCTs of measures such as different types of cash transfer. But economists worry that some RCTs might not be helping the people who are worst off. One reason for their concern is that many trials do not gather enough information to let researchers assess accurately whether they are actually reducing poverty. For

example, an RCT might show that an educational programme helps children to stay in school for longer, but that does not necessarily mean that poverty (and inequality) are being reduced. Unless scientists collect and report data on the socio-economic status of their participants before and after the trials, we simply don't know.

Things are starting to change: researchers are thinking about ways to improve the design of RCTs to account for equity and inclusion from the outset. A primary goal must be to ensure that researchers capture factors such as participants' ethnicity, culture and socio-economic background when they are relevant.

Real life is complex, and running trials that are large enough to achieve statistically valid observations about particular sub-populations is difficult. Scientists need to include people from poor and under-served populations as equal partners in studies, taking their perspectives into account from the design phase. Researchers must also collect and report data in ways that allow them to be aggregated in systematic reviews of the literature.

In the modern world, acute poverty and ill health exist alongside extreme wealth concentrated in the hands of relatively few people. This is morally reprehensible. In 2015, world leaders committed to eliminating poverty in all its forms by 2030, as one of the United Nations Sustainable Development Goals. Before the COVID-19 pandemic, the world had been making steady, albeit much too slow, progress on this front. But then the pandemic reversed the trend and sent millions back into extreme poverty. Today, the World Bank's best estimate is that at least 657 million people — 8% of the global population — survive on less than US\$1.90 per day. That is 36 million more than projections for where we would have been in 2020 if the pandemic hadn't happened, and 76 million more than such projections for 2022.

Ultimately, the world needs a new generation of technologies and policies that can both eliminate poverty and reduce inequality — while still protecting the environment. Economists and other social scientists need every tool in their bag as they seek to help policymakers to solve these grand challenges of our time. That will include RCTs — but to truly succeed, trials must have equity baked in.

*Nature* **606**, 624 (2022)

*doi:* <https://doi.org/10.1038/d41586-022-01675-2>

---

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-022-01675-2>

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

- WORLD VIEW
- 22 June 2022

# I have funds to buy reagents, but not remedies



Ignoring the challenges of research in low-income countries only perpetuates inequity.

- [Kondwani Jambo](#)

Recently, I was showing visitors our brand new, white-walled laboratories at the Malawi-Liverpool-Wellcome Programme (MLW) in Blantyre, Malawi. To demonstrate the power of our US\$250,000 flow cytometer, I used a \$250 vial of antibodies, obtained through grant funding from the UK Medical Research Council. I did this on the first floor, overlooking the city's Queen Elizabeth Central Hospital.

There, people are dying because they lack access to a \$2 course of generic antibiotics or to a hospital bed. Holding a reagent worth enough to treat

more than 100 people, as I look out of the window at a hospital with many more patients than the 1,350 it officially has space for, it's impossible to not feel guilty.

I am an immunologist at the MLW, where I lead a research group, and I'm also a tenured senior lecturer at the Liverpool School of Tropical Medicine, UK. My research focuses on understanding human immune responses, including the nature of herd immunity, with the aim of optimizing vaccinations.

I tell myself I could be elsewhere in the world, holding the same vial of antibodies — and the Queen Elizabeth Central Hospital would still be underfunded and overcrowded. But I'm in Blantyre, which means I'm constantly torn between my research and moral obligations. It's gut-wrenching to study antimicrobial resistance using expensive genomic sequencing platforms, while not being able to provide a generic antibiotic to a patient who walked kilometres to reach the hospital.

But that's just how the funding works. Like many global-health researchers, I am permitted to spend money on research consumables or kitting out my lab, but can't financially contribute to local clinical infrastructure such as hospitals or supplies of medicines. This feels unfair.

After navigating that moral maze, we also face practical considerations when it comes to doing research here in Malawi. Although my lab can afford expensive reagents, it takes at least two months for me to order and receive a specific antibody. I have to predict follow-up experiments with great accuracy to purchase reagents well ahead of time and in sufficient volumes. It becomes a major challenge when reviewers demand follow-up experiments, for little real value, without considering the work this would entail. The effect, albeit unintended, is to undermine diversity in science. It's more difficult for scientists here — where health research can have the most impact — to comply than it is for those at lavishly funded universities in Europe or North America, far away from anyone needing malaria treatment.

The solution is broader recognition, by both funders and journal publishers, of the challenges researchers like myself face.

Funders could consider partitioning existing global health grants so that some of the money goes directly towards the development and maintenance of clinical-care infrastructure in countries where research is taking place.

I've seen this happen already. After senior leaders at the London-based biomedical funding charity Wellcome visited in 2017, they committed money directly to the Queen Elizabeth Central Hospital to establish the eight-bed high-dependency respiratory unit, which became a lifesaver — literally — for people with severe COVID-19.

Publishers, meanwhile, could provide a dedicated space for authors to describe the limitations and restrictions under which research is conducted. This could be a short section, entitled 'context of the research', in a paper where the authors can detail what it took to do the work, and what is and isn't possible in that particular setting. For example, when I carry out bronchoscopy studies in Malawi, I must avoid risky repeated invasive sampling of patients, because there are few critical-care facilities to treat them if anything goes wrong. If journals allowed manuscripts to include such detail, then editors and reviewers could make better-informed decisions about the relative value of follow-up experiments, and focus instead on the work already generated and its potential benefit to human health.

Some might object to making special provisions for research in low-income settings. I've heard arguments that funding clinical care should not be part of support for research; that if it were, it would dilute funds or create ethical problems for health providers, who might feel that they must woo researchers to attract the money needed to help patients. And some tell me that descriptions of logistical difficulties do not belong in a research paper: that their inclusion could serve as an excuse for lack of rigour.

But these proposed changes support worthy goals: helping people and conducting studies in real-world circumstances. Appreciating what researchers go through to generate data need not lower the bar for rigour. Understanding the context of research can aid interpretation, reproducibility and applicability elsewhere. Funders often support research for which the impact on patients in the short term is uncertain; perhaps funders would get

more value for money if they also supported immediate clinical-care needs. The benefits outweigh the drawbacks.

As I look through the window of my lab, leaning against a machine that cost a quarter of a million dollars, it's disheartening to know that if nothing is done, the logistical and psychological burden of conducting excellent research here in Malawi will grow, and the inequity gap will widen. Simple changes here could matter a great deal.

*Nature* **606**, 625 (2022)

*doi:* <https://doi.org/10.1038/d41586-022-01677-0>

---

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-022-01677-0>

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

# How COVID has deepened inequality — in six stark graphics

Troubling data show how the pandemic has exacted an unequal toll, pushing tens of millions into poverty and having the greatest effects on already-disadvantaged groups.

By Saima May Sidik  
22<sup>nd</sup> June 2022



Credit: Ernesto Benavides/AFP/Getty

Credit: Ernesto Benavides/AFP/Getty

The shock waves of the COVID-19 pandemic have reverberated throughout the world. But the latest data reveal that the burden of COVID and its aftermath does not rest equally.

In six graphics, *Nature* details how the pandemic has worsened existing inequalities and exposed others in terms of income, health, safety and more.

## The global picture

Credit: Martin Harvey/Getty

Credit: Martin Harvey/Getty

The past two years have been particularly challenging for the world's poorest people, and this is just the beginning. By the end of this year, at least 75 million more people will have been pushed into poverty (living on less than US\$1.90 a day) than was expected before the pandemic. The war in Ukraine and rising inflation have exacerbated the effects of the pandemic, as prices for food, fuel and nearly everything else have skyrocketed.

Researchers at the World Bank have estimated how the number of people in poverty has changed over the past few years, assuming that everyone's income rose and fell in proportion with the variation in income of an average person in their country. 'Baseline' estimates assume that inflation affects all demographics equally, whereas the pessimistic projections reflect the expectation that inflation will affect the lowest earners most.

In 2022, up to 677 million people could be living in extreme poverty — almost 100 million more than without the combined crises of the pandemic, inflation and the war in Ukraine.

Almost 500 million of these people could be in sub-Saharan Africa.

Projections for regions such as East Asia and the Pacific are lower — although there are not enough data to analyse poverty in some areas, such as South Asia.

Recovery is projected to take longer in the Middle East and North Africa than in many other regions.

As households adapt to higher prices, the number of people living in poverty could go down. However, if food prices continue to rise, poverty might deepen before it improves.

Source: World Bank

---

Another data set from the World Bank reveals how the people with the lowest incomes have had the largest financial losses. Using changes in gross domestic product as a proxy, researchers estimated how much income people have lost and how much they've recovered, compared with what they would have earned had there been no pandemic. The results show that although people at all income levels lost money, the highest earners regained more than half their losses between 2020 and 2021, whereas the lowest earners have not recouped their losses compared with expected increases in earnings.

Researchers estimate that in 2021, the average incomes of people in the bottom 20% of earners were 6.7% lower than projected before the pandemic.

The situation is similar for earners in the second quintile.

Earners in the third quintile saw greater recovery, making only 5.1% less than they would have without the pandemic.

The incomes of earners in the fourth quintile are bouncing back even more quickly.

And the highest 20% of earners are doing best of all, making only 2.6% less than they would have without the pandemic.

Source: World Bank

---

Watch a [video version of this story](#).

## Uneven tolls

Credit: Richard Tsong-Taataarii/Star Tribune/Getty

Credit: Richard Tsong-Taataarii/Star Tribune/Getty

In addition to causing economic shocks, COVID-19 has had unequal effects on people's health. Data from the US Centers for Disease Control and Prevention show that death rates in the United States have been particularly high in Indigenous people: the rate is about 108% higher than that for white people and about 180% more than that for people of Asian descent. These data are age-adjusted, a statistical correction that allows for fair comparisons between demographics with different age distributions.

COVID-19 death rates vary by race and ethnicity in a US data set. The rate in Indigenous groups is double that for the white population, for instance.

Source: CDC

---

The state of an individual's health often correlates with factors such as employment and housing. People living in areas that lack basic necessities such as adequate housing or education often have worse health than do those living in regions that have these essentials. The same relationship holds true for COVID-19, as a UK analysis shows. Epidemiologist William Palmer at the UK health think tank Nuffield Trust in London used data from the Office for National Statistics and the Office for Health Improvement and Disparities to compare death rates from COVID-19 and other health conditions against the Indices of Multiple Deprivation, which measure how deprived an area is on the basis of factors such as average income, employment rate, education level and crime rate.

At the peak of the first wave of COVID-19, between March and July 2020, the death rate in areas of England that scored highest for deprivation was double that of the least-deprived regions. Similar degrees of disparity were seen for some other conditions.

In England in 2020, the COVID-19 death rate in the most deprived areas was double that in the most affluent.

Sources: UK Office for National Statistics; Office for Health Improvement and Disparities

## Social risk factors

Credit: Andre Coelho/Bloomberg/Getty

Credit: Andre Coelho/Bloomberg/Getty

Researchers in various countries have devised other ways to measure the risk of disease in different groups. Looking at Brazil, economist Luiza Nassif Pires at the Levy Economics Institute of Bard College in Annandale-on-Hudson, New York, and her colleagues assigned each state (indicated by a dot in the graphic) a social risk factor based on metrics such as how many people own cars — alleviating the need to use public transportation — and how many live in crowded conditions. Residents of states with higher risk factors were less able to socially distance themselves. Just before the peak of the pandemic's first wave in 2020. The team found that people in areas with higher risk factors contracted and died from COVID-19 at higher rates than did those in areas with lower scores.

The risk of catching and dying from COVID-19 correlates with social vulnerability in Brazil.

To calculate vulnerability, researchers looked at factors such as how readily people in different states could socially distance or whether they owned their own car.

Source: Levy Economics Institute of Bard College Annandale-on-Hudson

---

The pandemic has also had an unequal effect on other public-health problems, exacerbating violence against women, for instance.

Across 13 low- and middle-income countries, 45% of women surveyed by the United Nations Entity for Gender Equality and the Empowerment of Women (UN Women) say that they or a woman they know has experienced violence during the pandemic. Violence can be physical, such as hitting or throwing an object at someone; verbal, such as yelling; and it can involve denying basic needs such as food or health care; or involve sexual harassment.

This graphic shows the proportion of women who think that violence against women has increased in their community since the pandemic began. The researchers surveyed at least 1,200 women in each country using phone interviews. Because they were not able to travel or have direct contact while doing the surveys, the researchers cannot compare their results directly with studies done before the pandemic.

A survey of thousands of women in 13 lower- and middle-income countries suggests that violence against women increased during the pandemic.

Source: UN Women

---

---

## **SPRINGER NATURE**

Springer Nature

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

© 2022 Springer Nature Limited. All rights reserved.

[TopBuilt with Shorthand](#)

---

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-022-01647-6>

- NEWS FEATURE
- 22 June 2022

# These experiments could lift millions out of dire poverty

Randomized trials are changing the way governments and aid organizations study — and deliver — measures to reduce inequality and poverty.

- [Jeff Tollefson](#)



People taking part in a poverty-reduction trial in Niger receive cash and record-keeping booklets. Credit: Andrea Borgarello for World Bank

In 2012, the government of Niger began giving some of its poorest citizens free money. Over the next few years, around 100,000 participating households received 24 monthly payments of roughly US\$16 — which more than doubled their typical spending power.

The programme was based on decades of evidence from carefully controlled trials, suggesting that simple cash infusions can transform lives. And Niger is not alone: cash transfers have become a popular tool as governments try to alleviate poverty.

Several years in, the effort in Niger would also serve as a crucial testing ground for a new generation of expanded assistance programmes that offer people various types of personal, social and economic support in addition to hard cash. In a report issued last year<sup>1</sup>, the World Bank identified more than 200 such programmes in 75 countries, which collectively reach nearly 92 million people.

But that's just a fraction of the number of people living in extreme poverty. More than 650 million people across the globe get by on less than US\$1.90 per day, with severe impacts on public health and social and political stability in many areas. Many countries have also witnessed deepening inequality, a trend that has only increased since the COVID-19 pandemic began.

As in Niger, many of the latest anti-poverty programmes are grounded in science. Starting in the 1990s, researchers began to run randomized controlled trials — assigning participants to either receive an intervention or not — to test the effectiveness of various forms of help, ranging from subsidies for textbooks to direct distribution of money. Now, governments and aid organizations are starting to scale up the most promising strategies. They are also asking new questions about how to tackle inequality — and how to make sure programmes benefit those who need them most.

Some of the latest work comes from the Sahel region of Africa, where as much as 80% of the population lives in extreme poverty. Researchers from several leading non-governmental research organizations teamed up with the World Bank to work out ways to get more bang for each buck. Initial results from these trials, involving more than 50,000 households in Niger, Senegal,

Mauritania and Burkina Faso, are showing the benefits of add-ons to cash transfers, such as business training and coaching to promote self confidence<sup>2</sup>. And the World Bank is already working with governments in eight countries across the global south to scale up — and study — similar schemes, says Arianna Legovini, an economist who heads the bank's Development Impact Evaluation programme, headquartered in Washington DC.

Researchers recognize that meticulously planned and carefully delivered trials are not real life; nor are they easy to replicate across countries and cultures. One of the biggest challenges today, for instance, is to work out how to scale up pilot programmes to larger populations. Nonetheless, patterns are emerging: cash grants can be cheap and effective; early-childhood intervention can work; empowering women can have lasting impacts. Like many of her fellow scientists, Legovini is hopeful about the challenge of moving from science to public policy. “We’re not there yet,” Legovini says, “but I think we’re getting better.”

## Cash benefits

One of the primary lessons from rigorous research into poverty reduction should not, perhaps, come as a surprise: giving people money makes them less poor. “It’s almost arithmetic,” says David Evans, an economist with the Center for Global Development in Washington DC.

And yet, it took years of research to allay fears that poor people, given free money, would decline legitimate work and squander their new-found resources on temptations such as alcohol and tobacco<sup>3</sup>. Not only did the ‘labour/leisure trade-off’ fail to materialize, but once economists started running trials, the opposite effect prevailed time and again: free money translated into free time, and poor people tended to use that time productively.

In one trial<sup>4</sup> conducted in 2011–13, economists tested a simple cash-transfer programme in Kenya. Two groups of roughly 250 participants received the equivalent of around \$400, either as a lump sum or broken up into 9 equal payments; among those, 137 households received an extra \$1,100 over the

course of 7 months. The monthly instalments tended to promote food security, whereas lump-sum payments were more likely to be spent on durable goods such as furniture. In all cases, however, economic activity and psychological well-being improved. Subsequent work in Kenya<sup>5</sup>, involving the infusion of about US\$1,000 to more than 10,500 poor households across 653 randomized villages, suggested that the resulting increase in economic activity also benefited those who did not receive the payments.

“People are strapped, they are struggling, and when they get more money, that frees them up to do more, not less,” says Dean Karlan, co-director of the Global Poverty Research Lab at Northwestern University in Evanston, Illinois, and a principal investigator on the Sahel trials. “Pretty consistently, that’s what we see.”

Some programmes come with conditions intended to promote education or public health, but the easiest and cheapest to administer is the unconditional cash transfer. “That’s our benchmark,” says Legovini. “If you can’t do better than that, why bother?”



A villager who took part in a universal-income study feeds chickens at her home in Kenya. Credit: Yasuyoshi Chiba/AFP/Getty

What economists have found is that often, they can do better. In 2015, Karlan was part of a team that published a landmark study<sup>6</sup> documenting persistent gains, in terms of income as well as physical and mental health, from a programme that offered more than just money. Piloted in six countries on three continents, the intervention provided a productive asset such as goats or chickens, as well as temporary monetary support and longer-term educational resources such as entrepreneurial training and life-skills coaching. This multi-pronged ‘graduation’ approach has become the gold standard, but economists are still tinkering with the model.

In the latest iteration in the Sahel<sup>2</sup>, Karlan and his colleagues varied the interventions across more than 300 villages in which participants were already receiving small cash infusions through the government grant programme. Beginning in 2016, everybody participating in the treatment

groups received business training, including an entrepreneurial course designed for illiterate people and regular sessions facilitated by local coaches. They also formed their own savings-and-loan association, designed to promote cash savings and give out loans in times of need.

In addition to the training, one group received a cash grant worth around US\$310. And instead of cash, another group received exercises designed to promote interpersonal communication, community empowerment and other life skills. This second group also watched and discussed with other villagers a film about a couple who overcome various personal and economic challenges with the support of their family and community. The last treatment group received both the cash infusion and the extra psychosocial support.

Although communities that received the full treatment performed best, the results suggest that psychosocial interventions were as important as the cash infusion. Indeed, if cost effectiveness is the only consideration, the psychosocial intervention without the extra cash grant outperformed the other interventions. For Karlan, the message is clear. “We need to start thinking harder about some of these other aspects of poverty,” he says. “It’s not just about the money.”

## Scale matters

Scientists have learnt this lesson before. In a classic experiment<sup>7</sup> that predates the current wave, paediatrician Sally Grantham-McGregor tested in-home interventions designed to bolster nutrition, mother–child interactions and cognitive development among more than 125 malnourished Jamaican children, who were aged 9–24 months at the start of the 2-year study in 1987. Two decades later, children in the treatment group earnt 25% more than their untreated peers; after three decades<sup>8</sup>, the income disparity increased to 37%.

“When I started, I was told it was nonsense: you couldn’t work with these mothers, because they were not educated enough,” says Grantham-McGregor, who ran the experiment at the University of West Indies in

Kingston before moving to University College London. “Now it’s accepted that you can work with them, and you can have an impact.”

But Grantham-McGregor’s experience also demonstrates the fundamental challenges in running, interpreting and scaling up such experiments. Although she is amazed at the long-term effects of her study, she readily acknowledges that there is no way to determine what caused that impact: did the interventions work because they boosted the children’s cognitive development, or changed the mothers’ behaviour, or both?

And then there’s the challenge of expansion. When scientists tried to replicate the Jamaican experiment with around 700 children in Colombia and some 70,000 in Peru, the interventions had significantly smaller effects<sup>9</sup>. The lesson, Evans says, is both simple and daunting: scaling up interventions that depend on complex human interactions won’t be easy.

“The fundamental technology is very simple for cash-transfer programmes,” says Evans. “But with parent training, people are going in and building relationships and helping parents become better parents. That is more difficult.”

To the extent that they rely on personal and social interventions, Evans warns that more-complex programmes such as the one in Niger could face similar challenges as governments try expanding them in the future.

Although replicating the Jamaican experience has proved difficult, Grantham-McGregor says there is little doubt that early interventions can change the lives of disadvantaged children. Her former colleagues at the University of the West Indies have developed online materials for the programme, dubbed Reach Up, that have been used by government agencies in several countries. But, she says, it’s crucial to target interventions to the most disadvantaged people. “If you are interested in equity, if you can’t reach them all, for God’s sake reach the poorest children.”

## Missing targets

The need to ensure an even reach is on a lot of researchers' minds. As broader concerns about racism, equity and inclusion have entered the spotlight in recent years, many scientists have taken a second look at randomized trials focused on issues such as poverty and public health. They have realized that these trials can systematically exclude those who need help the most.

Lawrence Mbuagbaw, a health researcher at McMaster University in Hamilton, Canada, first encountered this problem in his home country of Cameroon a decade ago, when he was studying whether text messages could help to ensure that people with HIV took their medication on time. The messaging system worked, but phone ownership was skewed towards the wealthy — and in those families that shared a phone, towards men<sup>10</sup>. Mbuagbaw came to realize that public-health research often reflects inequity — whether in the form of persistent technological divides, language barriers or cultural and geographic divisions — instead of tackling it.



A group of women take part in an educational activity related to Niger's cash distribution study.Credit: Andrea Borgarello for World Bank

In Cameroon, Mbuagbaw says, most of the research is conducted in French — a minority language there — and in a single region around the capital. “All of our health policy is based on research conducted in one province,” he says.

And even in wealthy countries such as the United States, where most of the clinical research is conducted, many trials are run from hospitals in wealthy urban centres and consequently miss poor and often minority populations. If equity is a goal in such cases, Mbuagbaw says, then capturing a more representative population might require giving out travel subsidies or meal vouchers.

Other researchers share these concerns and are working on guidelines for designing and reporting better data from randomized trials. “Fundamentally, clinical trials are done for people who look very like me: middle-aged, affluent, fairly well-educated white males who are straight,” says Shaun Treweek, a health researcher at the University of Aberdeen, UK. He rattles off examples: COVID-19 vaccine trials that excluded pregnant people, trials focusing on cognitive behavioural therapies for depression in which 93% of the participants are white, trials of all sorts that exclude older people or are limited to English speakers for no apparent reason.

Treweek and his colleagues, through an initiative called Trial Forge, have been developing a framework for clinical-trial design that aims to boost the participation of people from minority ethnic groups and other under-represented populations. The framework was launched in October 2020 and, shortly thereafter, the UK National Institute for Health Research (NIHR) adopted it as formal guidance. The message is clear, Treweek says: if you don't address questions about potential bias and diversity in your grant applications, “you are probably dead in the water”.

Scientists won't always have the statistical power to assess how an intervention affects different groups of people, Mbuagbaw acknowledges. But at a minimum, they can help other researchers to do so. They can collect and report data on race and ethnicity, gender, education and socio-economic

status among their participants, he says, enabling other researchers to pool data from many studies in a systematic review.

Similar concerns have been raised about global development: research into poverty interventions is often absent in the poorest countries, owing to conflict and political instability. And where there are trials, it's not always clear that the interventions being tested actually benefit those most in need.

In some cases, it's the better-off people in a given treatment group who benefit the most, says Annette Brown, principal economist at FHI 360, a company based in Durham, North Carolina, that works on global-development issues. She cites a trial<sup>11</sup> that began in Tanzania in 2010, in which people were given cash grants that depended on school attendance; researchers found that students from households that were better off were more likely to complete primary school than were their poorer peers.

And all too often, randomized trials of poverty interventions do not actually report information about the current wealth of the recipients or whether the interventions actually reduced poverty, says Markus Goldstein, an economist with the World Bank in Washington DC. A trial of an agricultural intervention, for instance, might focus on technology adoption while ignoring one of the most important questions: did its adoption promote wealth?

“We could do a better job of describing, at the very least, the levels of poverty of the people in our studies,” says Goldstein. Karlan says no trial is perfect, and one focused on ultra-poor people could also miss the broader population of poor people. “You just have to be clear about the research you are doing,” he says.

The challenges facing researchers and governments trying to tackle poverty and inequality will only increase. As efforts to scale up promising interventions continue, Legovini says, scientists must also seek to understand the long-term impacts on poor people, as well as the broader economic impact on communities. The good news is that science is no longer an afterthought in the development world. After starting the World Bank’s Development Impact Evaluation programme with no staff in 2009,

Legovini now has a small army of roughly 250 people, including consultants.

Moving forward, she hopes the same scientific tools will expand to cover other development programmes, including those targeted at infrastructure, climate and governance. It won't happen quickly, but Legovini thinks science can help policymakers to make better decisions across the board. "Development is complex, and we need to be committed to working in different settings and building knowledge and understanding of how to get things right."

*Nature* **606**, 640–642 (2022)

doi: <https://doi.org/10.1038/d41586-022-01679-y>

## References

1. Andrews, C. *et al.* *The State of Economic Inclusion Report 2021* (World Bank, 2021).
2. Bossuroy, T. *et al.* *Nature* **605**, 291–297 (2022).
3. Evans, D. K. & Popova, A. *Econ. Dev. Cult. Change* **65**, 189–221 (2017).
4. Haushofer, J. & Shapiro, J. *Q. J. Econ.* **131**, 1973–2042 (2016).
5. Egger, D., Haushofer, J., Miguel, E., Niehaus, P. & Walker, M. W. Working Paper 26600 (NBER, 2019).
6. Banerjee, A. *et al.* *Science* **348**, 1260799 (2015).
7. Grantham-McGregor, S. M., Powell, C. A., Walker, S. P. & Himes, J. H. *Lancet* **338**, 1–5 (1991).
8. Gertler, P. *et al.* Working Paper 29292 (NBER, 2021).

9. Araujo, M. C., Rubio-Codina, M. & Schady, N. Working Paper Series IDB-WP-1230 (IDB, 2021).
  10. Mbuagbaw, L. *et al.* *PLoS ONE* 7, e46909 (2012).
  11. Evans, D. K., Gale, C. & Kosec, K. Working Paper 563 (CGD, 2021).
- 

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-022-01679-y>.

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

- BOOK REVIEW
- 22 June 2022

# Tackling inequality takes social reform

In separate books, leading economists explore the wide-ranging changes needed to produce a more just society.

- [Richard Wilkinson](#) &
- [Kate Pickett](#)



The COVID-19 pandemic has been hardest for underprivileged people, such as these children studying in an improvised classroom. Credit: Anindito Mukherjee/Getty

**A Brief History of Equality** *Thomas Piketty* Belknap (2022)

**The Origins and Dynamics of Inequality: Sex, Politics, and Ideology** *Jon D. Wisman* Oxford Univ. Press (2022)

The cold war rivalry between capitalism and communism ended not because the problems of capitalism had been solved, but because communism had failed as a solution. Still, the rivalry was good for capitalism. From the end of the First World War until around 1980, differences in the incomes of people in rich countries shrank. Welfare systems became increasingly generous and capitalism developed a more human face. But when the Soviet Union and Eastern European countries ceased to be an economic threat to Western capitalism, the rival system that had made capitalism more compassionate disappeared, allowing market systems to become harsher. Top tax rates for high earners were reduced, trade unions were weakened and income gaps widened.

Income inequality continues to trend upwards. Since 1995, almost 20 times as much of the increase in global wealth has gone to the richest 1% of people as to the poorest 50%. The global charity Oxfam estimates that 8 men now own the same amount of wealth as the world's poorest 3.6 billion people. But do these grotesque levels of inequality really matter? And if they do, who do they harm, and why?

Two books attempt to tackle these questions and lay out plans for far-reaching reform. The first comes from French economist Thomas Piketty, and the second from US economist Jon Wisman. Both argue that making the world a better place requires a reduction in inequality, and that inequality persists because of ideological beliefs, including that social position reflects innate ability, that the highest incomes are a payment for talent and that the economic system is too delicate to be tampered with. But there are important differences in what these authors think drives inequality and on what can be done to reduce it.

In *The Origins and Dynamics of Inequality*, Wisman takes seriously both the Marxist criticism of the market and the failure of communism. But his solutions to inequality are partly shaped by his view that economic inequality is driven by sexual competition and the aphrodisiac properties of

status, wealth and power. He quotes Saint Augustine on the sins of lust for money, power and sex, and says it wasn't until Charles Darwin that it became clear that the first two are driven by the third.

This perspective leads Wisman to argue that inequality can be reduced by taming the forces of sexual competition or, at least, by uncoupling them from money and power. In his view, the 'mating game' should instead be about gaining recognition for achievements in fields such as poetry, science, art or sport, or even through generosity or environmental credentials. He quotes US psychologist Geoffrey Miller, who says that finding better ways to manage human sexual competitiveness should be at the explicit core of social policy.

## Routes to economic equality

Piketty is also progressive — and has a stature approaching that of a latter-day John Maynard Keynes. His widely acclaimed *Capital in the Twenty-First Century*, published in 2013, demonstrated how economic inequality had historically arisen because the rate of return on capital, from which rich individuals derive their wealth, normally exceeds the economic growth rates on which most people's incomes depend.

In *A Brief History of Equality*, Piketty shows that greater equality is part of the long arc of historical progress. Between 1780 and 2020, he says, most regions and societies shifted towards greater equality, if not of income, then of rights and recognition — expansion of the right to vote, equality before the law, gender and racial equality, women's property rights and more.

Research by others has shown how new institutions and social, economic and political norms move together; for example, societies with smaller overall income differences, such as Scandinavian countries, also tend to have smaller gender differences in pay and political representation. This is not enough, however — societies still need to tackle inequities in people's abilities to flourish, for example in education or governance.

At heart, both these authors approach their subject from an emphatically economic perspective. Both accept the market as a necessity but want to remove its most inegalitarian and antisocial results by changing the context

in which it operates. Rather than doing away with private ownership of the productive system, both want to democratize it with forms of employee ownership and autonomy. Capitalism, they each argue, can be transformed by much more progressive taxation, more generous welfare and expanding forms of employee ownership of companies. Piketty explains that the twentieth century demonstrated that having “almost confiscatory tax rates” for the highest incomes — at times exceeding 80% in the United Kingdom and the United States — contributed to the long decline in inequality before 1980. Policies such as these could, as Wisman says, “eliminate the core source of exploitation that Marx identified within capitalism”.

## Harms of inequality

Why do both authors see the degree of inequality as the defining issue of history and as the challenge of our times? There is a striking absence of discussion, in both books, of the harm that inequality does. Perhaps this is a reflection of how, as an academic discipline, economics has failed to see the key psychosocial processes through which inequality makes itself felt. By contrast, a large body of research from other disciplines, including our own field of epidemiology, shows that inequality needs to be understood in more than monetary and structural terms.

Inequality needs to be seen as a social relationship. It places us in a hierarchy, ranked one above another, and — crucially — determines the social distance between us. Instead of encouraging the public spiritedness, cohesion and trust that can flourish in a community of near-equals, big material differences make class and status more important, exacerbating feelings of superiority and inferiority. As a result, people become more conscious of their status. The social structure ossifies and social mobility declines. In short, inequality is a social stressor.

That explains why more-unequal societies have worse physical and mental health, more antisocial behaviour — including higher homicide rates and more people in prison — and lower levels of child well-being and development. Inequality is not just an economic condition; it gets under our skin and into our minds, shapes our behaviours and fundamentally

undermines our collective well-being and flourishing. It causes chronic stress.

Similarly, primatologists have shown that subordinate status is damaging to the health of monkeys; it would be unethical to reproduce these experiments with human participants, but the findings mirror observations relating to people with a low social status. Piketty and Wisman travel further than most beyond the narrow boundaries of mainstream economic thinking, but the discipline needs to grapple with the psychological and sociological implications of the subject of inequality to truly understand the interplay of social and economic forces.

Although they are missing some of the human costs of inequality, Piketty and Wisman do agree on one wider harm: inequality greatly impairs our ability to minimize climate change and the cascade of problems it entails. Not only does it intensify status-related consumption, it also increases the political power of wealthy individuals, who cause the most environmental pollution while being affected by it the least.

Piketty ends by showing that tackling inequality is crucial to the kinds of power politics that will make or break efforts to tackle the climate emergency and other environmental crises. He foresees that political hostility will increase towards the high-income countries and individuals most responsible for the environmental crisis. In particular, he points to the likely shift of prestige and influence from the United States to China if the former ceases to be the standard bearer for democracy and the latter is able to remind the world that, despite human-rights abuses, it bears little historical responsibility for carbon dioxide emissions, slavery or colonialism. Piketty predicts that to limit the growing influence of China's authoritarian socialism, Western powers must abandon their hypercapitalist ideology and transition to a participative market socialism that is post-colonial, responsive to low- and middle-income countries, and able to respond effectively to the environmental crisis.

So inequality matters. It matters for people, and for the planet. It matters for all of us, and not just those at the sharp end of poverty and deprivation. Inequality is, as world leaders and thinkers such as Piketty and Wisman point out, a defining challenge of these times. If we want more than just a

more equitable distribution of resources, if we want sustainable prosperity for the world, we ignore these important and readable books at our peril.

*Nature* **606**, 643-644 (2022)

*doi:* <https://doi.org/10.1038/d41586-022-01680-5>

---

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-022-01680-5>

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

- COMMENT
- 22 June 2022

## Not all inequalities are alike

Better data and new statistical techniques could enable researchers to measure the form of inequality that seems most harmful to society — inequality of opportunity.

- [Francisco H. G. Ferreira](#)



A man picks through a rubbish bin outside a gated house in a wealthy suburb of Johannesburg, South Africa, in 2019. Credit: Waldo Swiegers/Bloomberg via Getty

At the onset of the COVID-19 pandemic, with job losses affecting 17% of all UK workers by early April 2020, women in the United Kingdom were 4.8 percentage points more likely than men to have lost their jobs<sup>1</sup>. In South Africa in 2017, the average income for a household of white people (adjusted for size and composition of households) was 5.6 times that for a household of African people ('African' is a recognized racial classification in South Africa<sup>2</sup>). And in the United States, a 2018 report found that people born in the 1980s to parents who were in the bottom half of the 'years of schooling' distribution had only a 13% chance of making it to the top quartile of the schooling distribution in their generation<sup>3</sup>.

In all of these cases, the gaps between the groups are not defined by how hard people work or study, by how much they save or by how responsible they have been. Instead, the gaps are solely down to characteristics over which individuals have no control: gender, race and upbringing, respectively. These are examples of what economists call inequality of opportunity.

Various lines of evidence show that inequality of opportunity is the component of inequality that is least acceptable and most harmful to society. Yet too few economists or other researchers try to quantify its contribution in their measures of a population's inequality in education, income, employment, health or other outcome of interest.

Doing so requires as much data as possible about people's circumstances — their race or ethnicity, the neighbourhood they grew up in, their parents' level of education and so on. Even when such data can be obtained, there isn't widespread agreement on how best to quantify the share of inequality in a population that is due to inequality of opportunity. This needs to change.

Given the importance of inequality of opportunity in public discourse and in emerging economic research, governments and researchers should be collecting the data needed to construct credible estimates of it. And economists and others should be developing better tools and methods to make those estimates.

## Circumstances matter

The ideal of equal opportunity has featured in Western political discourse since the Enlightenment in the seventeenth and eighteenth centuries, and remains prominent to this day.

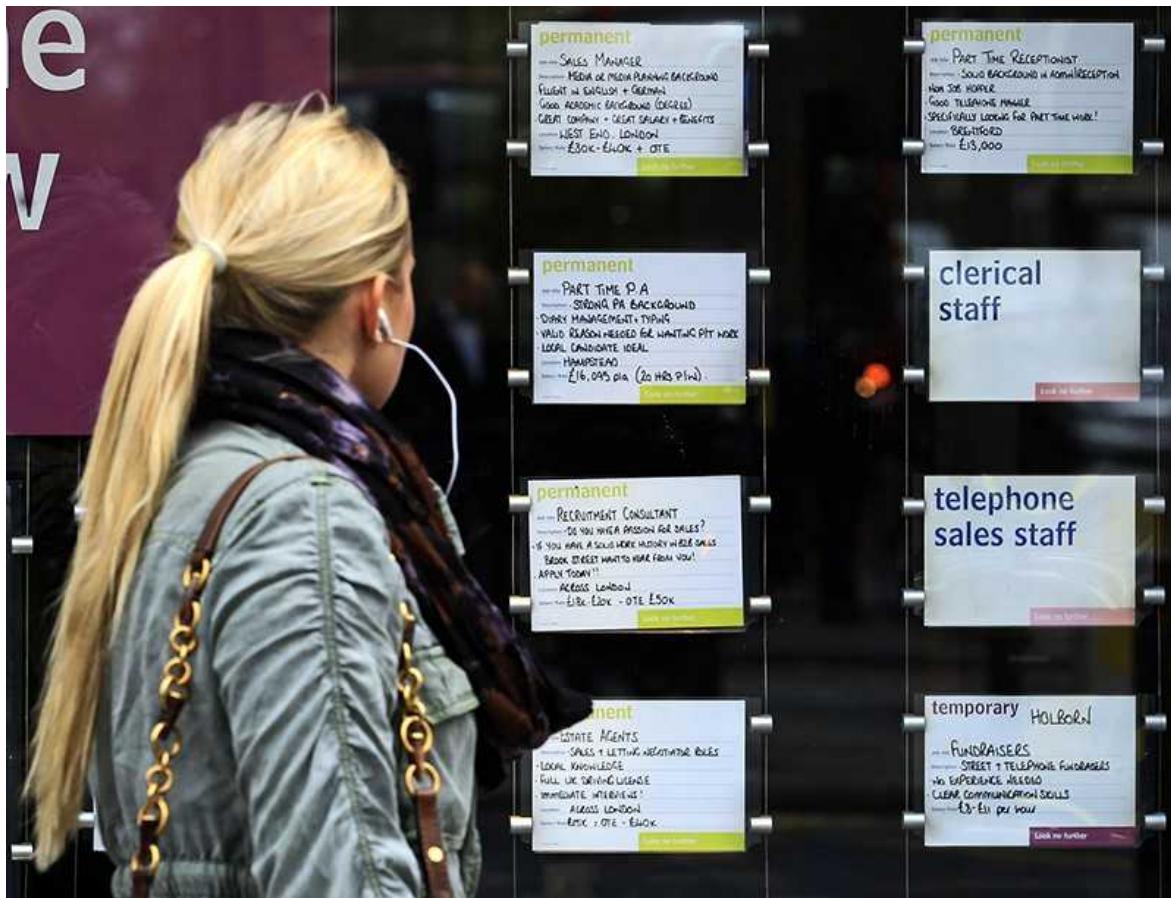
In his 10 June 1936 address at Little Rock, Arkansas, for example, US President Franklin Roosevelt said, “We know that equality of individual ability has never existed and never will, but we do insist that equality of opportunity still must be sought.” Likewise, the US philosopher John Rawls, in his highly influential 1971 treatise *A Theory of Justice*, reasoned that fair equality of opportunity — the idea that everyone in society should have the same access to goods, services and employment opportunities — is one of two principles of social justice. (The other is that all citizens should have the highest amount of freedom consistent with everyone else.)

Yet the concept remained intangible until economists, including Marc Fleurbaey, John Roemer and Dirk Van de gaer, attempted to formalize these ideas in the 1990s<sup>4–6</sup>. They suggested that all determinants of a particular social outcome (such as earnings) could, in principle, be divided into factors that individuals can and cannot control. They defined the first category as a person’s efforts, and the second as circumstances. This simple formulation provided a basis for an economic theory of inequality of opportunity.

Subsequent work — mainly in behavioural economics — has shown that it’s not just philosophers and economists who care about the distinction between circumstances and efforts. In multiple laboratory and field experiments since the early 2000s, individuals have been given real money and told to distribute it in whatever way they wish — such as between themselves and another participant, or among a group of fellow participants.

In such situations, only a few people make choices in line with those expected from ‘*Homo economicus*’ — a figurative species that is characterized, among other things, by purely self-interested preferences. When a participant is given US\$100, for instance, they are more likely to give away a sizeable chunk (\$30 or \$50) instead of giving away \$1 and keeping \$99. What’s more, if the participant distributing the money proposes an allocation that is deemed too unfair, potential recipients often give up the chance of a real monetary payment to punish the distributor<sup>7</sup>.

In short, humans (and even some animals) demonstrate an intrinsic preference for fairness and equity.



Early in the COVID-19 pandemic, UK women were more likely than men to have lost their jobs. Credit: Luke MacGregor/Reuters

When economists dug deeper into these behaviours, it became evident that the type of inequality people find most objectionable is that arising from factors seemingly beyond an individual's control.

In a 2010 study, for example, economists hired 238 students to type as many words as possible during a 10- or 30-minute period<sup>8</sup>. Each student could choose either the short or the longer task. They were paid according to the number of words they got right, but, at random, some were paid \$0.08 for each correct word, others \$0.16.

When the students were subsequently organized into pairs, and each member of a pair was told the working time, word output and payment of the other, each student could propose some redistribution of earnings within the pair. Most chose to compensate participants who had been randomly allocated a low wage. Fewer chose to compensate for the duration or quality of work, both of which were seen as being within the individual's control<sup>8</sup>.

## A social bad

There is also growing evidence that inequality of opportunity is especially detrimental to other social outcomes, such as economic growth.

Since the 1990s, researchers have investigated associations between income inequality and factors such as levels of violent crime<sup>9</sup> or disease and death<sup>10</sup>. This work has generally found that inequality leads to worse health and social outcomes. But the evidence for the relationship between inequality and growth has been much less conclusive, with different data sets and methods leading to disparate results<sup>11,12</sup>.

In the past decade, associations between growth and inequality have become clearer, thanks to researchers separating out the share of inequality that stems from inequality of opportunity<sup>13,14</sup>.

One study, for instance, found that there was no statistically significant association between inequality and economic growth across 26 US states between 1970 and 2000 when total income inequality was considered. But once total inequality was broken down into one component that was due to predetermined circumstances (inequality of opportunity) and another that was due to other factors, the former did have a significant negative impact on growth<sup>13</sup>.

Other work supports the idea that when sizeable groups are denied access to productive opportunities — simply on the basis of their personal circumstances — the resulting waste of human potential leads to lower economic efficiency and dynamism. A 2019 study, for example, examined the proportion of white men in the population of doctors and lawyers in the United States — a figure that fell from 94% in 1960 to 62% in 2010. The

work found that this decrease (and similar declines in other professions) could account for 20–40% of the growth in gross domestic product per person over that 50-year period. The authors suggest that these gains could have arisen from a more efficient allocation of talent, as women and Black men (the two other groups studied) increasingly took professional opportunities that had previously been closed to them<sup>15</sup>.

But if inequality of opportunity really does constitute the ‘active ingredient’ of inequality — the part that is most widely disliked, and which accounts for its most pernicious social effects — why are measures of the concept still relatively rare in public debate?

## Data deficit

‘Opportunity’ refers to the set of possibilities that are open to a person. As such, it is a difficult concept to quantify. Yet, building on the work of Roemer, Van de gaer and Fleurbaey, economists have proposed relatively simple ways to measure inequality of opportunity.

Essentially, these seek to quantify how much of the inequality in an outcome of interest (in income or education, say) is due to circumstances over which people have no control, while accounting for the fact that a person’s efforts are themselves influenced by their circumstances<sup>16–18</sup>.

Obtaining such a measure requires accurate information on as many predetermined circumstance variables for each person in the sample or population as possible. Data on race or ethnicity, sex or gender and place of birth are relatively straightforward to acquire. But ideally, quantifying inequality of opportunity also requires detailed, long-term information on a person’s family history — including the education, income and occupations of their parents, and even parenting behaviours.



Surgeons Robert Bloodwell, Denton Cooley and Grady Hallman (from left) in the 1960s, when some 90% of US doctors and lawyers were white men. Credit: AP/Shutterstock

Such data are rare in low- and middle-income countries, and even in some affluent ones. They do exist, however, for some high-income countries, such as the United States, Germany, the United Kingdom and most of Scandinavia.

Ultimately, two gold-standard kinds of data should be collected and made available in all countries for researchers and policymakers to gain a better understanding of the unequal opportunities people face across the world.

The first comes from detailed, longitudinal household surveys. Examples include the US Panel Study of Income Dynamics, which began in 1968, and Germany's Socio-Economic Panel, which has run since 1984. Such panels provide rich information on the parents and parenting behaviours of today's adults, and allow economists to link information across generations. Similar initiatives have begun in emerging economies, such as Mexico and Indonesia, but they are still very much the exception.

The second kind comes from administrative data sets that connect personal identifiers across generations, and across different aspects of people's lives: educational outcomes can be linked to employment and health histories, social-security contributions, tax payments and so on.

Providing researchers with access to such rich and sensitive data brings challenges around privacy and confidentiality. But these data are available to researchers in countries such as Denmark and Norway — subject to procedural constraints designed to protect people's privacy. Similar advances are happening elsewhere, including in Chile.

In short, with sufficient investment, it is not hard to imagine how the collection of data on people's circumstances could be bolstered worldwide.

## Statistical conundrum

Even with the best possible data, economists can accurately assess inequality of opportunity only if they work out precisely how variables should be used to partition a population into groups with homogeneous circumstances.

In theory, when data are available for an entire population, all the circumstance variables in those data should be used. Even then, some variables will go unobserved. Researchers might have information, for instance, on whose parents went to university, but not on the rankings of the universities. Because some circumstances will always be overlooked, any estimate of inequality of opportunity will be an underestimate.

On the flip side, when data are available only for a sample of the population (as is typically the case), using a large number of variables to partition the sample can lead to over-fitting. Ever-finer subgroups have ever-fewer people in them, causing group-specific estimates to be noisy. This sampling error tends to inflate the share of variation that is explained by inequality of opportunity. Given these two contradictory biases, how can researchers decide how best to partition a population into subgroups by circumstance variables?

One promising approach is to use machine-learning techniques<sup>19,20</sup>. Given a data set with a certain number of circumstance variables and subcategories ('race', for example, could be divided into 'Black', 'white' and 'Asian'), an algorithm tries every possible way of splitting the sample into two groups, and makes the split that generates the most significant statistical difference between the two means. This process is then repeated for each subgroup until no further significant differences are found (for some standard level of statistical significance agreed in advance).

In an unpublished analysis, my colleagues and I applied this method to a sample of just over 6,000 South African households, whose income and other data had been collected by the country's National Income Dynamics Study in 2017. Our sample included people who self-identified as one of four groups: African, Coloured (a recognized racial classification in South Africa), Asian/Indian, or white.

We instructed an algorithm to split this sample, leaving at least 100 observations in each final group, and used 1% as the cut-off for statistical significance between two groups. Given these parameters — and using only predetermined circumstances such as parental education and parental occupation to make each split — the algorithm divided the sample into ten subgroups or types.

Next, we gave every person in each subgroup an income equal to that subgroup's average, and computed inequality over this 'smoothed' distribution, so that only inequality of opportunity (between the subgroups) is left.

Inequality of opportunity estimated in this way is between 66% and 74% of the overall income inequality in South Africa, as measured by a Gini coefficient of 0.61. (The Gini coefficient measures inequality in the distribution of household income, with lower values indicating a more equal distribution.) Using our lower estimate, 66% of 0.61 is 0.40, which is similar to the latest Gini coefficient reported by the World Bank for the United States (0.42 in 2019; see [go.nature.com/3xpjmwp](https://go.nature.com/3xpjmwp)). In other words, the inequality between those ten subgroups of the South African population is essentially the same as the overall inequality in all of the United States. Using a more conservative measure (the mean logarithmic deviation), the

share ranges from 39% to 48% of the total. For comparison, a 2015 study of South Africa using 54 types found that inequality of opportunity accounted for only 17–24% of the mean logarithmic deviation<sup>21</sup>.

This machine-learning approach must be tested in many settings to assess whether it truly represents the extent of inequality of opportunity in a society. Some might worry that certain algorithms are methodological ‘black boxes’. But in a context where no causal hypothesis is being tested, and researchers are seeking only the most efficient way of using the available data to measure inequality of opportunity, a machine-learning approach seems difficult to beat.

## One and all

Inequality of opportunity is the conduit through which inequality is reproduced between generations. Combined with improved data-collection efforts, new computational statistical approaches could help to quantify its true extent in regions and countries around the world.

Such information would complement measures of economic mobility (the association between an adult’s income and that of their parents), which rely on a single circumstance variable: parental income. It could also complement measures of overall inequality in outcomes, such as earnings, wealth or education.

Indeed, opportunities and outcomes are best seen as two sides of the same coin: a family’s outcomes today help to shape the opportunities of their children tomorrow, and those opportunities, in turn, help to determine those children’s future outcomes.

*Nature* **606**, 646–649 (2022)

doi: <https://doi.org/10.1038/d41586-022-01682-3>

## References

1. Adams-Prassl, A., Boneva, T., Golin, M. & Rauh, C. *J. Public Econ.* **189**, 104245 (2020).
2. Brunori, P., Ferreira, F. H. G. & Peragine, V. in *Prioritarianism in Practice* (eds Adler, M. D. & Norheim, O. F.) Ch. 11 (Cambridge Univ. Press, 2022).
3. Narayan, A. *et al.* *Fair Progress? Economic Mobility Across Generations Around the World* (World Bank, 2018).
4. Fleurbaey, M. *Theory Decis.* **36**, 277–307 (1994).
5. Roemer, J. E. *Equality of Opportunity* (Harvard Univ. Press, 1998).
6. Van de gaer, D. *Equality of Opportunity and Investment in Human Capital*. PhD thesis, Katholieke Univ. Leuven (KU Leuven) (1993).
7. Henrich, J. *et al.* (eds) *Foundations of Human Sociality: Economic Experiments and Ethnographic Evidence from Fifteen Small-scale Societies* (Oxford Univ. Press, 2004).
8. Cappelen, A. W., Sorensen, E. Ø & Tungodden, B. *Eur. Econ. Rev.* **54**, 429–441 (2010).
9. Fajnzylber, P., Lederman, D. & Loayza, N. *J. Law Econ.* **45**, 1–39 (2002).
10. Marmot, M. G. *et al.* *Lancet* **337**, 1387–1393 (1991).
11. Banerjee, A. & Duflo, E. *J. Econ. Growth* **8**, 267–299 (2003).
12. Ferreira, F. H. G., Lakner, C., Lugo, M. A. & Özler, B. *Rev. Income Wealth* **64**, 800–827 (2018).
13. Marrero, G. A. & Rodríguez, J. G. *J. Dev. Econ.* **104**, 107–122 (2013).
14. Bradbury, K. & Triest, R. K. *Russell Sage Found. J. Soc. Sci.* **2**, 178–201 (2016).

15. Hsieh, C.-T., Hurst, E., Jones, C. I. & Klenow, P. J. *Econometrica* **87**, 1439–1474 (2019).
16. Bourguignon, F., Ferreira, F. H. G. & Menéndez, M. *Rev. Income Wealth* **53**, 585–618 (2007).
17. Checchi, D. & Peragine, V. *J. Econ. Inequal.* **8**, 429–450 (2010).
18. Ferreira, F. H. G. & Gignoux, J. *Rev. Income Wealth* **57**, 622–657 (2011).
19. Hothorn, T., Hornik, K. & Zeileis, A. *J. Comput. Graph. Stat.* **15**, 651–674 (2006).
20. Brunori, P., Hufe, P. & Mahler, D. G. *The Roots of Inequality: Estimating Inequality of Opportunity from Regression Trees* Policy Research Working Paper No. 8349 (World Bank, 2018).
21. Piraino, P. *World Dev.* **67**, 396–405 (2015).

---

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-022-01682-3>

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

- CAREER FEATURE
- 22 June 2022

# The rise of inequality research: can spanning disciplines help tackle injustice?

Wide-ranging expertise and direct involvement of those affected will help to make inequality research more meaningful.

- [Virginia Gewin](#)

[Find a new job](#)



Sasha Henriques worries that the data that guide genetic counsellors are not representative.Credit: Ekow Oliver

Vida Maralani began her career studying whether education is, in fact, the ticket to socio-economic advancement and reducing inequalities, as she'd been taught. "I've evaluated some of the most expensive social policies our government has ever funded to prevent school drop-outs," says Maralani, interim director of the Center for the Study of Inequality at Cornell University in Ithaca, New York. Yet she found that policies targeted at children in low-income families who were at risk of dropping out of school could only do so much. "These kids were not moving out of their neighbourhoods, and their mums had no more resources than they did before," she says. "The disadvantages that surrounded their community did not change."

Generally, the unequal or unjust distribution of resources and opportunities in a society is studied in just one dimension, such as through income or education, says Maralani. Yet inequalities in income, wealth, education, health and access to technology are inter-related and differ by gender, race , ethnicity and geographical location in important ways. The root causes are multidimensional and dynamic. Some of the most influential work of the past decade — notably French economist Thomas Piketty's 2013 book *Capital in the Twenty-First Century* — demonstrated [how persistent inequality has become](#), even raising international concern.

There's an urgency driving increased interest in inequality research. "The reason for it is horrific — inequality is growing," says Melanie Smallman at University College London, who studies how technology contributes to inequality. More than two-thirds of countries are experiencing increased income disparity, which exacerbates the risks of divisions and hampers economic and social development, according to the *United Nations World Social Report 2020* (see [go.nature.com/3hpyjkw](https://go.nature.com/3hpyjkw)).

COVID-19 has only raised the bar for inclusive research, because scientists can now assess the impact of the pandemic on those inequalities. Income losses due to the pandemic are estimated to be higher in people paid the least — pushing many deeper into poverty and jeopardizing the prospects of future generations, according to the World Bank.



Vida Maralani says there's no simple fix for inequalities driven by complex factors. Credit: Robert Barker

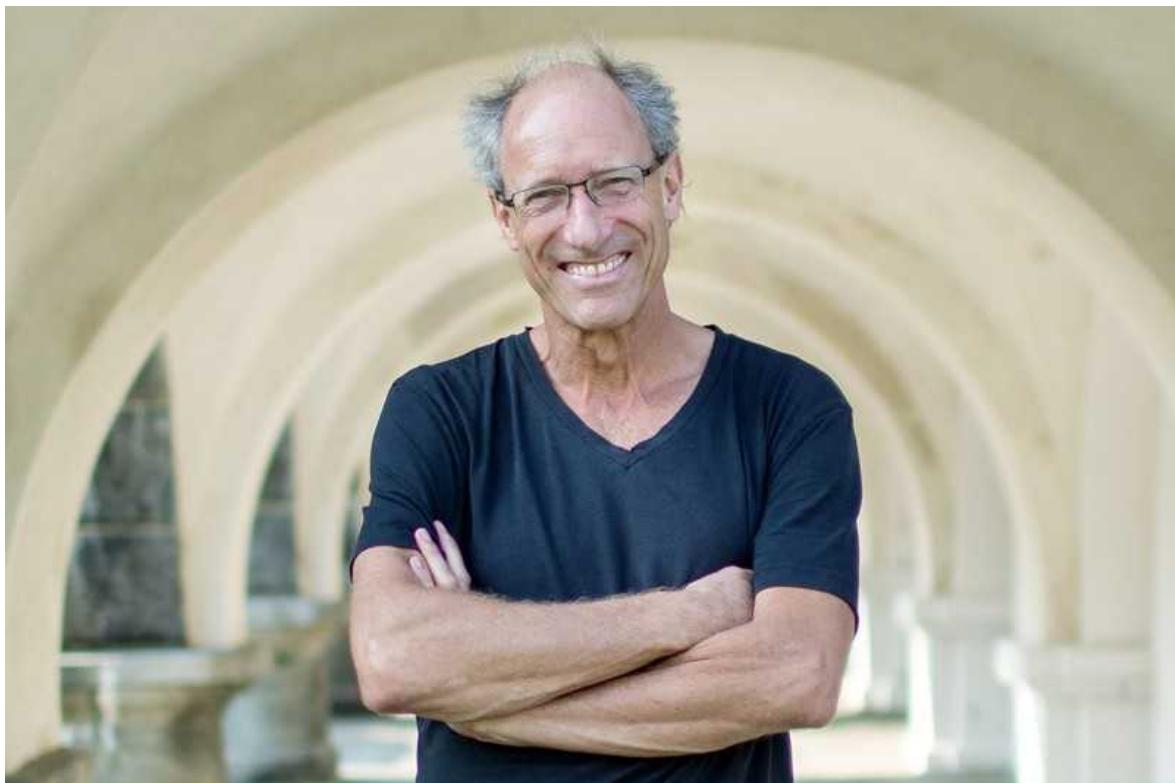
Not surprisingly, the field is expanding. “The concern that worsening inequality will be long-lasting is one of the key reasons those in the field expand their research often via challenging interdisciplinary approaches — and increasingly through a lens of justice,” says Maralani.

As researchers attempt to disentangle the complex drivers of worsening inequality, the expansion of the sphere of scholars and organizations exploring inequality “is a silver lining”, says Maralani. A lot of people with the right skills are needed to collaborate across fields as disparate as sociology, health and climate change. The challenge, she says, is to go beyond well-worn paths of how we think about inequality and think more holistically.

Take human migration. Demography has long been studied by sociologists, but migration is now influenced heavily by natural resources and climate change. “As we get more data, our ability to study things really expands,”

Maralani says. “And as we are better able to share and visualize and describe our data, then our engagement with the public expands.”

Technological change, the decline of trade unions, low minimum wages and the proliferation of no-contract gig jobs are all contributing factors in the continued increase in inequality in the United States since 1980, raising concerns among researchers, policymakers and politicians. One of the UN’s Sustainable Development Goals is to [reduce inequality in and between countries](#). And numerous funders — including the Ford Foundation, the Russell Sage Foundation and the William T. Grant Foundation, all in New York City, and the MacArthur Foundation in Chicago, Illinois — have launched initiatives to “reduce inequality”, and called for researchers to develop evidence-based strategies to meet the goal.



Don Kalb warns that the green technology transition could increase inequalities.Credit: Eivind Senneset

“There’s a broad feeling that this monster is too big for any one academic profession to claim as the job,” says Don Kalb, a social anthropologist at the University of Bergen in Norway. There is also a push to monitor technology

shifts in real time to try to prevent inequality from getting even worse. The move away from fossil fuels is a prime example. “The green transition has a very high probability of increasing all sorts of inequalities”, in part because most green-technology leaders are in high-income countries, he says. Researchers need to assess the political tools that might be needed to prevent that source of inequality — particularly given that society now has the capacity to capture real-time data relating to that goal. If successful, the research could even make a leap towards social justice.

## Bringing society into research

One goal of inequality research is to make the production of science itself more egalitarian. As interdisciplinary approaches swell to meet growing demand from funding agencies and journals, scientists are grappling with [how to meaningfully involve members of the public](#).

During the ten years that Sasha Henriques was a genetic counsellor in the United Kingdom and South Africa, she had nagging concerns about her work. She wondered whether the data being used to counsel patients about genetic conditions were adequately representative. And the lack of diversity among counsellors bothered her. To explore the intersection between race, ethnicity, ancestry and genomics, Henriques has started a PhD at the University of Cambridge, UK. “It’s all intersectional,” she says, “nobody belongs to just one group.”

Henriques’s research will identify when it is relevant to include race in research data, how best to categorize human populations beyond race and ethnicity and how to make the benefits of genomics and health research equitable. But the issues are so complex that Henriques made sure her work was guided by her core value: offering people the tools and information they need to help them understand their genetic risk for disease and make any necessary health and lifestyle decisions — options that have been less available to historically excluded populations.

She advises other inequality researchers, especially those wanting to break into the field, to do the same. “It can feel quite overwhelming to narrow

down the research to do something meaningful without going down all sorts of rabbit holes,” says Henriques.

Another key strategy is making sure that oppressed communities are directly involved in the core research design. For example, some health-equity researchers caution that publishing work on historically excluded groups that were not involved in the process might lead to specious conclusions that only perpetuate bias<sup>1</sup>.

If the goal is to find realistic solutions to inequality, Henriques notes, it is important to bring society into the scientific process. “It can be intentionally embedded within research and research design,” she adds. As Henriques works out how to do that in her research, she has created a website to begin conversations with members of the public (see [www.geneticsengage.org](http://www.geneticsengage.org)).

Bringing people into research prevents the production of science from becoming yet another site of inequality. It also helps to unearth researchers’ hidden biases and ensures greater confidence in the research outcomes. “It matters who produces the knowledge,” says Susanne Koch, a sociologist at the Technical University of Munich in Germany who studies how inequalities shape forest and environmental research.



Susanne Koch encourages inequality researchers to dismantle structural dynamics that can be discriminatory. Credit: Nela Dorner

In May, Koch ran a workshop with environmental researchers from all over the world to discuss the structural issues that can lead to discriminatory research practices. “We need to produce diverse knowledge that is often not put on the table because of social inequalities in academia,” she says. For example, in international forest ecology research, which has conventionally focused on natural resources, it can be “hard for researchers to acknowledge that their knowledge, as is everyone’s, is rooted in their unique contexts”. It struck Koch that colleagues from developing nations said the workshop was the first time they had a chance to address issues that had long bothered them.

Koch says if researchers just produce statistics that show how bad a situation is and then walk away, what incentive do the affected communities have to embrace their findings? “Making this second step after producing the statistics is often missed and what I think is necessary — and, of course, it is difficult.”

## A trendy warning

Given the burgeoning interest in inequalities, seasoned researchers worry that newcomers, sometimes viewed as interlopers, will simply tack equity onto projects because it is trendy, and thus end up publishing sub-par analyses. Elle Lett, a social epidemiologist at the University of Pennsylvania in Philadelphia, and her colleagues have referred to such researchers as “health equity tourists”, who are “at risk of polluting the health equity landscape and riddling the academic record with ineffectual, and potentially harmful studies that mischaracterize root causes of health inequities and obfuscate potential solutions”<sup>1</sup>.

Steven Roberts, a racial-inequality researcher at Stanford University in California, says that adding an inequality dimension simply to be “trendy” can foster a careless, even reckless, approach that does not engage with issues as deeply and as meaningfully as it should. For example, he says, research focused on a white population might just state that future work will need to factor in diversity, yet offer little description of why that is necessary, what it would look like and why people of colour weren’t included in the first place.

Roberts is finishing a sabbatical year at the Russell Sage Foundation, where he is the only psychologist in a cohort of economists, sociologists, political scientists and population-studies researchers who are sharing insights on how their fields investigate inequality. Essentially, each person presents their project — ranging from defining white privilege to identifying racial inequality in scientific publications — and receives crowdsourced feedback across disciplines. “It’s been the best interdisciplinary experience of my life, and it’s been a rare thing,” he says. Still, he realizes it is difficult to maintain that level of cross-fertilization, especially as an early-career researcher.



Steven Roberts cautions against research that just pays lip service to inequality. Credit: Stanford University

And not every facet of the research publication pipeline is equipped to embrace this level of interdisciplinary collaboration. “The more interdisciplinary you are, the better your understanding of the issue and the scope of it, but in terms of communicating that to a non-interdisciplinary audience, it gets tricky,” Roberts says. “My experience is that journals have their lens on their discipline, and you fit into that framework, or you don’t get published,” he says. He is embarking on a project to document publication barriers faced by scientists of colour across the sciences.

Still, like many, Roberts sees inequality research becoming more solutions-focused and thinks it will energize research that transcends disciplines. He says that when researchers come together to bring their own discipline’s perspectives and methods to the table, “we are in a great position to tackle big problems”.

## Data and technology challenges

One of the biggest hurdles for researchers studying inequality in communities is a lack of data. Bonnielin Swenor, director of the Johns Hopkins University Disability Health Research Center in Baltimore, Maryland, who is visually impaired, studies the discrimination and oppression that affects the disabled community — the biggest minority group in the United States, at 67 million adults. In July 2020, Swenor co-authored a plea in *The Lancet Public Health* to include disability information alongside age, race, ethnicity and gender identity when gathering data about COVID-19 — and to make collection of such data routine to create a more equitable society<sup>2</sup>. Incomplete data, she argued, would result in unreliable disease risk estimates.

With US\$1.5 million in financial support from Johns Hopkins, Swenor started a research centre in 2019 to address such data oppression. The centre collects data on disability disparities in employment, education and even COVID-19 vaccine access, and aims to share those data to maximize health and equity for people with all types of disability. “Without the data, it’s as if inequities don’t exist,” she adds.



Bonnielin Swenor argues for more disability data sharing to maximize equity. Credit: Will Kirk

Smallman agrees. In the wake of COVID-19, she and her colleagues started the [UK Pandemic Ethics Accelerator](#) to look retrospectively at the role of data collection in making the effects of COVID-19 better or worse. She notes that 40,000 older people died in care homes in England in the first year of the pandemic, yet these numbers weren't released until many months later.

Swenor, like many, distinguishes between inequality and inequity. Inequities — unfairness or bias in a community — leads to inequalities. And that framework goes one step further. "Justice is addressing the reasons why we don't have equality," she explains. Swenor says there hasn't been enough focus on the perspectives of people with disabilities and with other intersecting marginalized and oppressed identities. "If you're only going to try to solve the equation of inequality with one variable, you're never going to solve that equation."

Smallman studies how technology such as artificial intelligence exacerbates inequalities. From driverless cars to robots that harvest crops, many artificial-intelligence innovations aim to replace humans. In doing so, Smallman explains, the wealth generated will be held by a handful of people rather than shared as occurred with past innovations. "The business model of advanced digital technology companies is to drive more inequality," she says. She advocates for scientists to engage in ethical discussions about how technologies affect individuals and communities, shape institutions and even affect the planet and future generations. "We need to try to stretch out our ethical thinking, so that we can begin to account for these really profound effects that advanced technologies are having," she says.



As inequalities grow globally, so does interest in research in causes and solutions, says Melanie Smallman. Credit: Falling Walls Foundation

One advantage, however, is that improvements in computational power will allow researchers to comb through decades of qualitative data, such as ethnographies and case studies, to gain fresh insights. “We’ll be able to look across case studies and identify patterns which we weren’t able to see before,” she says. And that will be important globally, because inequality is not going to go away, she says, adding that climate change will rapidly accelerate it.

Trying to understand the social determinants of inequality is really humbling, says Maralani. The factors underlying inequalities are dynamic and hard to measure, change over the life course and are passed down through generations. For Kalb, these challenges demand robust scholarship from both veteran and newcomer researchers alike. “We need to brainstorm together,” he says.

To be effective, inequalities researchers will need to move beyond data collection and analysis, says Swenor. They must “dismantle the root cause of

the sources of inequities — the policies, the institutions and the systemic issues that create inequities for certain groups,” she says. “That is the future.”

*Nature* **606**, 827-829 (2022)

*doi:* <https://doi.org/10.1038/d41586-022-01684-1>

## References

1. Lett, E. *et al.* *J. Med. Sys.* **46**, 17 (2022).
2. Reed, N. S., Meeks, L. M. & Swenor, B. K. *Lancet Public Health* **5**, E423 (2020).

---

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-022-01684-1>

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

- WHERE I WORK
- 22 June 2022

# Shark researcher attacks lack of diversity in marine science

Jasmin Graham has a passion for predators and promoting inclusive science.

- [Virginia Gewin](#)



Jasmin Graham is the president and chief executive of Minorities in Shark Sciences in Bradenton, Florida. Credit: Cassie Wegeng for *Nature*

I fell in love with the ocean growing up fishing with my family. Once I learnt you could get paid to study fish, I sought out a marine-science degree at the College of Charleston in South Carolina, where a mentor's enthusiasm for sharks rubbed off on me. I then went on to study another predator — sawfish — for my master's degree at Florida State University in Tallahassee.

Predators keep fish numbers on a reef in balance, which prevents overgrazing and disease spread. Sharks have been on the planet longer than trees or dinosaurs, and I study how the millions of years of evolution have led to the weird faces of animals such as hammerhead sharks and sawfish. I also conduct surveys of predators to document them in the wild. In this photo, we are about to release a blacknose shark (*Carcharhinus acronotus*). In just a couple of minutes, we had measured it, taken a fin sample for genetic analysis and a muscle biopsy to look at mercury levels and tagged it for identification.

Another part of my job at Mote Marine Laboratory & Aquarium in Sarasota, Florida, is to increase diversity in marine science. I train researchers to be allies of interns from minoritized groups by being mindful of how identities affect peoples' experiences.

As president and chief executive of Minorities in Shark Science, a non-profit organization that I co-founded in 2020 and that has 489 members from 33 countries, I instruct institutions and individuals on how to overturn oppressive systems. We focus on removing barriers — often financial — that exclude women of colour from marine sciences. We provide funded field experiences, professional development and diversity training.

Organizations still do a lot of box-ticking to add 'diverse' people rather than addressing root causes of inequality. My colleagues and I highlight the importance of a strong code of conduct, accountability and training. Having an army of allies drowns out the bad players.

*Nature* **606**, 832 (2022)

doi: <https://doi.org/10.1038/d41586-022-01685-0>

---

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-022-01685-0>

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

# This Week

- **UK's rupture with Horizon Europe is totally unnecessary**

[ 21 Jun 2022]

- Few ever expected British and EU scientists to be forced apart. Researchers must never take international collaboration for granted.

- EDITORIAL
- 21 June 2022

# UK's rupture with Horizon Europe is totally unnecessary

Few ever expected British and EU scientists to be forced apart. Researchers must never take international collaboration for granted.



EU research commissioner Mariya Gabriel has suggested the United Kingdom look at alternatives to Horizon Europe that could allow UK–EU collaborations to continue. Credit: Thierry Monasse/Anadolu Agency/Getty

When Britain left the European Union at the end of January 2020, researchers were assured that this did not mean leaving the EU's research programme, Horizon Europe. Under the terms of the United Kingdom's EU exit, the country would keep paying into the €95.5-billion (US\$100.6-billion) fund and researchers would continue to be able to access grants (including prestigious European Research Council (ERC) grants), lead projects and participate in initiatives such as the International Thermonuclear Experimental Reactor facility in France. Scientists let out heavy sighs of relief. Although most had strongly opposed Brexit, access meant that long-standing research partnerships would continue and new ones could be forged.

But a lot has changed since then. Relations between UK and EU policymakers have nose-dived, with researchers trapped in the middle. Those awarded ERC and other grants are now expected to lose them. The principal reason is the British government's decision to [break some of the terms of the separation agreement](#) that it carefully negotiated with the EU.

The UK government has introduced draft legislation into its parliament that is intended to amend trading arrangements between Northern Ireland (which is part of the United Kingdom) and the independent Republic of Ireland (which is a member of the EU). It is doing this unilaterally, instead of using the official dispute-resolution system. This action has triggered legal action by the EU against the United Kingdom for breaking international law.

While all of this is happening, the EU has halted research cooperation. UK recipients of EU grants have been told they will need to move to an EU institution if they want guaranteed access to the funds. Some are reluctantly preparing to do so. The EU's legal action is likely to make any future UK access to Horizon Europe much more difficult. The legal case will probably take several years to run its course, and Horizon Europe is time-limited: it ends in 2027.

Research leaders in both the EU and the United Kingdom have fought a vocal and high-profile campaign [called 'Stick to Science'](#), urging politicians to keep politics out of science. But, barring a last-minute change of heart, a science relationship that has lasted some five decades looks likely to come to an end. If and when that happens, it could be the biggest setback to

European science cooperation ever seen. Over the years, researchers in mainland Europe have enriched UK science no end — and vice versa.

Unsurprisingly, relations between the UK government and the nation's scientists are at one of their lowest points in recent memory. Researchers are exasperated over the uncertainty and the lack of detailed communication about what will come next, and have concerns about inconsistencies in the government's thinking on funding.

UK science minister George Freeman, a biotechnology entrepreneur and intellectual, is preparing a backup global fund for UK researchers that he is informally calling Plan B. Last week, Freeman told a parliamentary inquiry that the government will publish a 'prospectus' for this fund before Members of Parliament go on their summer break on 21 July. He added that the fund will include international fellowships for UK researchers and more funding for high-risk, high-reward science similar to that funded by the US Defense Advanced Research Projects Agency.

One problem for the minister is that the UK Treasury — the department that is providing the funding — needs to know which of the two options to fund. If the country won't be joining Horizon Europe and Plan B isn't ready in time, there's a fear that some of the allocated funds could be diverted to other spending priorities.

Another reason the scientific community has little confidence in Britain's funding ambitions is the government's decision to abruptly end one of the nation's existing (and popular) global funding schemes, the Global Challenges Research Fund (GCRF), along with the decision not to renew a second global fund, the Newton Fund, when it ended in 2021. The unexpected cancellation of the GCRF, in particular, created chaos for existing projects.

It is imperative that the UK government consults with some of the country's experts in research funding on the design of a replacement global fund. Consultation should also include organizations such as the Royal Society, the British Academy and the Royal Academy of Engineering, which were among those responsible for managing and disbursing the GCRF and the Newton Fund.

These funds supported partnerships between researchers in the United Kingdom and international counterparts, including many in low- and middle-income countries, particularly on projects aimed at meeting the United Nations Sustainable Development Goals. The funds transformed research at many universities, both in the United Kingdom and around the world. By 2019, the GCRF was supporting nearly 5,000 researchers working across more than 800 projects in some 120 countries. An evaluation of the lessons learnt from these experiences could be of huge benefit to the designers of the new global fund.

The story of the United Kingdom's scientific decoupling from the EU must stand as a warning to researchers around the world: international cooperation in science cannot be taken for granted. Researchers have come to expect that those elected to lead will understand that science and knowledge thrive on partnerships and international exchange — and that in times of political tension, disagreement or conflict, research, knowledge and scholarship should continue in spite of those differences. But the way that the United Kingdom's rupture with the EU has spilt into science shows that this is not necessarily the case.

As the world enters a much more uncertain phase following the pandemic and Russia's invasion of Ukraine, we urge all researchers to redouble their efforts to maintain and boost collaborations. No action is too small. Added together, acts of solidarity keep collaborations alive in the absence of formal ties, just as they did in previous times of tension and conflict.

*Nature* **606**, 623-624 (2022)

*doi:* <https://doi.org/10.1038/d41586-022-01676-1>

---

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-022-01676-1>

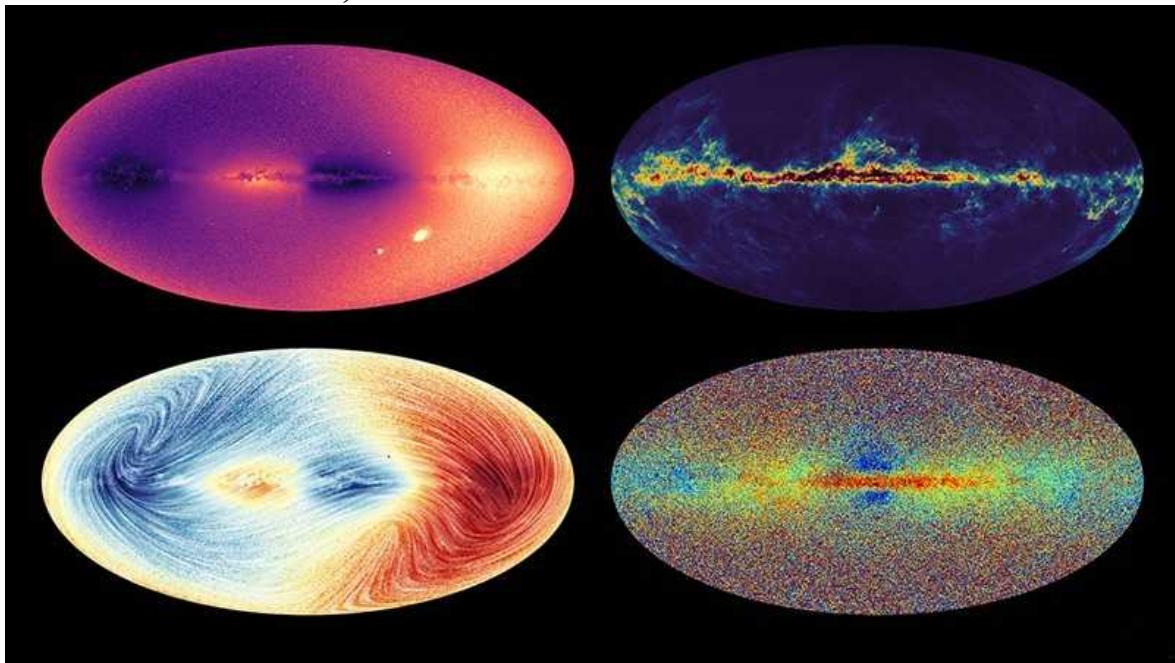
# News in Focus

- **[Milky Way mega-map and UK science turmoil](#)** [ 22 Jun 2022]
  - The latest science news, in brief.
- **[COVID and smell loss: answers begin to emerge](#)** [ 09 Jun 2022]
  - Researchers are learning more about how the SARS-CoV-2 coronavirus stifles smell — and how they might revive it.
- **[Max Planck's cherished autonomy questioned following criticism of misconduct investigations](#)** [ 08 Jun 2022]
  - Demoted archaeologist Nicole Boivin calls on the German government to oversee the prestigious research society — several other directors say that investigations into them lacked transparency.
- **[Legendary bacterial evolution experiment enters new era](#)** [ 14 Jun 2022]
  - A laboratory has been growing 12 populations of E. coli since 1988 — this year, the cultures will get a new custodian.
- **[Ancient DNA traces origin of Black Death](#)** [ 15 Jun 2022]
  - Genomes show that plague-causing bacteria found in Kyrgyzstan graves are direct ancestors of those that triggered the medieval pandemic.
- **[African researchers lead campaign for equity in global collaborations](#)** [ 10 Jun 2022]
  - Cape Town statement on research partnerships between the global north and south will highlight unethical practices and offer advice to scientists.
- **[New COVID drugs face delays as trials grow more difficult](#)** [ 13 Jun 2022]
  - Fewer people are eligible for the massive studies needed to test treatments for severe COVID-19.

- NEWS ROUND-UP
- 22 June 2022

# Milky Way mega-map and UK science turmoil

The latest science news, in brief.



The Milky Way in four maps: data from the Gaia spacecraft show the speed at which stars move towards or away from us, known as radial velocity (top left); their radial velocity and proper motion, or how they move across the sky (bottom left); their chemical make-up (bottom right); and the interstellar dust (top right). Credit: ESA/Gaia/DPAC/CU6 ([CC BY-SA 3.0 IGO](#))

**Mega-map of Milky Way adds depth to stars' motions**

Astronomers' main reference guide to the Milky Way has [received a major update](#). The Gaia mission, in which a spacecraft is tracking nearly two billion stars, has released a vastly improved map. The map now includes the 3D motions of tens of millions of stars and thousands of asteroids — as well as the detections of stellar ‘quakes’ and of possible extrasolar planets.

The mission's team unveiled the trove, which consists of 34 months' worth of data, on 13 June.

Gaia was launched by the European Space Agency in 2013 and orbits the Sun at a fixed distance from Earth. It takes repeated measurements of the same stars from different perspectives. This makes each star's apparent position change by a tiny angle — typically millionths of a degree — proportional to its distance. The mission team uses such changes and a technique called parallax to calculate the star's distance from the Sun.

The biggest addition to the previous catalogue is the set of detailed spectra for about one million stars. By measuring a spectrum's Doppler shift, the team has calculated 30 million ‘radial velocity’ measurements. Each indicates the speed at which a star is moving towards, or away from, the Sun. Together with Gaia's measurements of the star's motion across the sky and of its distance, the data provide a full reconstruction of the star's trajectory as it circles the Galaxy.



UK science minister George Freeman said that negotiations on the United Kingdom's involvement in Horizon Europe are at a "crunch point". Credit: Alamy

## Fears rise that Britain will leave EU research fund

Concerns are growing that the United Kingdom is on the [cusp of leaving the European Union's Horizon Europe research programme](#) after UK science minister George Freeman (pictured) said on 8 June that "time is closing" for a positive resolution.

With a budget of almost €100 billion (US\$106 billion), Horizon Europe provides research funding for scientists in EU member states and other nations that choose to become 'associate members'. But the United Kingdom's participation in the programme has been in question since the country voted to leave the EU in 2016. In December 2020, the EU and the United Kingdom struck an agreement as part of the overall Brexit deal to

continue the Horizon Europe collaboration. But issues concerning the deal's 'Northern Ireland protocol' have stalled its ratification.

UK participation in Horizon is "being used as a bargaining chip in a much broader and bigger political negotiation", says Kieron Flanagan, a science-policy researcher at the University of Manchester, UK.

The UK government has said that if an agreement cannot be reached, it will develop its own £15-billion (US\$18.7-billion) research programme to rival Horizon Europe.

*Nature* **606**, 629 (2022)

*doi:* <https://doi.org/10.1038/d41586-022-01678-z>

---

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-022-01678-z>

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

- NEWS
- 09 June 2022

# COVID and smell loss: answers begin to emerge

Researchers are learning more about how the SARS-CoV-2 coronavirus stifles smell — and how they might revive it.

- [Michael Marshall](#)



After a COVID-19 infection, some people have recovered their ability to perceive odours thanks to smell training. Credit: Fabio Bucciarelli/The New York Times/eyevine

Researchers are finally making headway in understanding how the SARS-CoV-2 coronavirus causes loss of smell. And a multitude of potential treatments to tackle the condition are undergoing clinical trials, including steroids and blood plasma.

Once a tell-tale sign of COVID-19, smell disruption is becoming less common as the virus evolves. “Our inboxes are not as flooded as they used to be,” says Valentina Parma, a psychologist at the Monell Chemical Senses Center in Philadelphia, Pennsylvania, who helped field desperate inquiries from patients throughout the first two years of the pandemic.

A study published last month<sup>1</sup> surveyed 616,318 people in the United States who have had COVID-19. It found that, compared with those who had been infected with the original virus, people who had contracted the Alpha variant — the first variant of concern to arise — were 50% as likely to have chemosensory disruption. This probability fell to 44% for the later Delta variant, and to 17% for the latest variant, Omicron.

But the news is not all good: a significant portion of people infected early in the pandemic still experience chemosensory effects. A 2021 study<sup>2</sup> followed 100 people who had had mild cases of COVID-19 and 100 people who repeatedly tested negative. More than a year after their infections, 46% of those who had had COVID-19 still had smell problems; by contrast, just 10% of the control group had developed some smell loss, but for other reasons. Furthermore, 7% of those who had been infected still had total smell loss, or ‘anosmia’, at the end of the year. Given that more than 500 million cases of COVID-19 have been confirmed worldwide, tens of millions of people probably have lingering smell problems.

For these people, help can’t come soon enough. Simple activities such as tasting food or smelling flowers are now “really emotionally distressing”, Parma says.

## Scrambled nuclei

A clearer picture of how SARS-CoV-2 causes this disruption should help to create better therapies for the condition. Early in the pandemic, a study

showed<sup>3</sup> that the virus attacks cells in the nose, called sustentacular cells, that provide nutrients and support to odour-sensing neurons.

Since then, clues have emerged about what happens to the olfactory neurons after infection. Researchers including biochemist Stavros Lomvardas at Columbia University in New York City examined people who had died from COVID-19 and found that, although their neurons were intact, they had fewer membrane-embedded receptors for detecting odour molecules than usual<sup>4</sup>.

This was because the neurons' nuclei had been scrambled. Normally, the chromosomes in these nuclei are organized into two compartments — a structure that enables the neurons to express specific odour receptors at high levels. But when the team looked at the autopsied neurons, "the nuclear architecture was unrecognizable," Lomvardas says.

Other studies suggest why only some people experience long-term smell loss. In January, a research team reported<sup>5</sup> finding a genetic mutation in people that was associated with a greater propensity for smell or taste loss. The mutation — a change to a single 'letter', or base, of DNA — was found in two overlapping genes, called *UGT2A1* and *UGT2A2*. Both encode proteins that remove odour molecules from the nostrils after they have been detected. But it's not yet clear how SARS-CoV-2 interacts with these genes.

There is also evidence of lasting changes to the brain for people with smell loss. In a study published in March<sup>6</sup>, 785 people in the United Kingdom had their brains scanned twice. About 400 people became infected with COVID-19 between scans, so the scientists were able to observe structural changes. The COVID-19 survivors showed multiple changes, including markers of tissue damage in areas linked to the brain's olfactory centre. It's not clear why this was the case, but one possibility is lack of input. "When we cut off input from the nose, the brain atrophies," says Danielle Reed, a geneticist also at Monell. "It's one of the clearest things we know about taste and smell."

## Treatments in testing

In the meantime, many treatments are being explored, often in small clinical trials. But it's still early days, so the only thing that most researchers recommend for now is smell training<sup>7</sup>. Patients are given samples of strong-smelling substances to sniff and try to identify, with the aim of driving the restoration of olfactory signalling. However, the method seems to work only with people who have partial smell loss, Reed says. That means it helps about one-third of people who experienced a chemosensory disruption after COVID-19, adds Parma.

To find treatments for everyone else, many researchers are exploring steroids, which reduce inflammation. COVID-19 is known to trigger extensive inflammation, which might play a part in smell disruption. So, in theory, steroids could help — but, in practice, the results have been disappointing. For instance, a 2021 study<sup>8</sup> gave smell training to 100 people with post-COVID anosmia. Fifty of them also received a nasal spray with the steroid mometasone furoate, while the other 50 did not. There was no significant difference in outcome between the two groups.

Another therapeutic possibility is platelet-rich plasma; this is made from patients' own blood and is rich in biochemicals that might induce healing. A pilot study published in 2020<sup>9</sup> followed seven patients who had platelet-rich plasma injected into their noses: five showed improvement after three months. Similarly, a preprint published in February this year<sup>10</sup> followed 56 people and found that platelet-rich plasma made them more sensitive to smells. But these are "really small numbers", says Carl Philpott, a nose and sinus specialist at the University of East Anglia, Norwich, UK. A US-based team is now launching a larger study.

Unlike COVID-19 vaccines, which were tested at unprecedented speed because of tremendous government support, treatments for post-COVID chemosensory dysfunction are plodding along. Philpott is in the early stages of a small study of vitamin A, which previous experiments have suggested could help with other forms of smell loss. "The reality is that the study will take the rest of this year to run, and it'll take us probably to the middle of next year before we analyse the data and report it," Philpott says. "If we find a positive benefit, our next job will be to apply for more funding to do a full stage trial."

*Nature* **606**, 631–632 (2022)

doi: <https://doi.org/10.1038/d41586-022-01589-z>

## References

1. Coelho, D. H., Reiter, E. R., French, E. & Costanzo, R. M. *Otolaryngol. Head Neck Surg.* <https://doi.org/10.1177/01945998221097656> (2022).
2. Boscolo-Rizzo, P. *et al.* *Rhinology* **59**, 517–527 (2021).
3. Brann, D. H. *et al.* *Sci. Adv.* **6**, eabc5801 (2020).
4. Zazhytska, M. *et al.* *Cell* **185**, 1052–1064.e12 (2022).
5. Shelton, J. F. *et al.* *Nature Genet.* **54**, 121–124 (2022).
6. Douaud, G. *et al.* *Nature* **604**, 697–707 (2022).
7. Addison, A. B. *et al.* *J. Allergy Clin. Immunol.* **147**, 1704–1719 (2021).
8. Abdelalim, A. A. *et al.* *Amer. J. Otolaryngol.* **42**, 102884 (2021).
9. Yan, C. H., Mundy D. C. & Patel, Z. M. *Laryngoscope Investig. Otolaryngol.* **5**, 187–193 (2020).
10. Steffens, Y. *et al.* Preprint at medRxiv  
<https://doi.org/10.1101/2022.02.14.22270109> (2022).

---

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-022-01589-z>

- NEWS
- 08 June 2022

# Max Planck's cherished autonomy questioned following criticism of misconduct investigations

Demoted archaeologist Nicole Boivin calls on the German government to oversee the prestigious research society — several other directors say that investigations into them lacked transparency.

- [Alison Abbott](#)



Archaeologist Nicole Boivin was a director at the Max Planck Institute for the Science of Human History in Jena, Germany. Credit: Saverio Petraglia

A former director of one of the Max Planck Society's prestigious research institutes, who says she was unfairly demoted, has called on Germany's research ministry to oversee the society's procedures for misconduct investigations. Six other Max Planck Institute (MPI) directors, some of whom have themselves been investigated or demoted for misconduct, have also told *Nature* that they feel the society's misconduct investigations lack transparency and are affected by bias.

Archaeologist Nicole Boivin at the Max Planck Institute for the Science of Human History in Jena is one of eight MPI directors who are known to have been demoted, or threatened with demotion, after investigations into allegations of non-scientific misconduct, which includes actions such as bullying and harassment. In an [open letter](#) published on 8 June, Boivin says that these investigations have been “plagued by allegations of bias, conflicts of interest, and procedural and legal shortcomings”.

The run of demotions has led to an atmosphere of fear among MPI directors, says developmental biologist Herbert Jäckle, an emeritus director at the Max Planck Institute for Multidisciplinary Sciences in Göttingen. “They are concerned about how the investigations are going, but afraid to speak out,” he says.

Boivin writes in her letter that she was not given the opportunity “to offer any reasonable response, evidence or witness testimony” to anonymous allegations.

The Max Planck Society (MPS) declined to answer specific questions for this article, but in an e-mailed summary of Boivin’s case, a spokesperson told *Nature* that the MPS stands by all of its decisions in the affair and that Boivin was given all required opportunities to present her side of the case. MPS president Martin Stratmann, who made the initial decision to demote Boivin, declined to comment on the case because it has concluded. But the MPS’s e-mail said that Stratmann’s decision “was preceded by a extremely thorough internal investigation into the allegations levelled against Dr Boivin”.

## Researchers speak out

Stratmann demoted Boivin last October after an investigation that stretched over more than two years concluded that she had committed misconduct, including bullying two young scientists and taking over another scientist's research project. In December, a Berlin court suspended the demotion, but the society reinstated it in March. Boivin, who remains a researcher at the institute, denies all the allegations against her. Last April, her PhD students and postdocs wrote to the MPS president in her defence.

Boivin's open letter joins criticisms made by many other scientists of how the society handles misconduct allegations. Last November, [145 high-profile female international researchers wrote an open letter](#) to the MPS leadership expressing concern that female scientific leaders at MPIs are being disproportionately affected by challenges to their leadership styles (the MPS has previously rejected charges of gender bias made in that letter). At around the same time, a personal letter from 24 emeritus institute directors to Stratmann, which was leaked to the press, questioned whether Boivin's investigation had appropriately heard her side of the story. It also said that the case threatened the society's reputation and called for the MPS to introduce transparent governance structures.

The MPS declined to give *Nature* official figures on the numbers of cases involving non-scientific misconduct at MPIs. But the public criticisms have fuelled a debate about whether Germany's publicly funded research organizations should have so much autonomy.

The MPS has an annual budget of around €2 billion (US\$2 billion) from state and federal governments to run its 86 institutes and facilities. It enjoys the freedom granted in the German constitution to organize its own structures and procedures, without political interference.

This freedom has sometimes been a source of friction in political quarters. "This extensive autonomy is desirable for designing and carrying out research, but should not necessarily extend to aspects of personnel," says Holger Becker, a physicist who is a lawmaker in the German parliament and is on the parliament's research committee. He says that the MPS has a very

strong staff hierarchy, and the president has an unusual amount of power compared with in research organizations in other countries.

## Dismayed members

Boivin's case began in 2018, when she made an official complaint to Stratmann that she was being harassed by the two other directors at her institute. She charged that Stratmann had failed to address her allegations seriously. Two weeks later, Stratmann informed Boivin that she was to be investigated for misconduct; Boivin says that the details of those accusations were not made clear to her at the time.

The society established a committee to investigate the allegations made by and against Boivin. In January 2021, the committee concluded that Boivin had engaged in both scientific and non-scientific misconduct. Vice-president Ulman Lindenberger investigated further, and reported on the scientific-misconduct allegations. On the basis of those two reports and the advice of his executive committee, Stratmann decided in October to immediately demote Boivin — without waiting for the approval of the MPS senate. The senate comprises MPS directors and representatives from politics and industry and notionally oversees the workings of the society. MPS regulations allow such a step only when there is a risk of immediate damage to the society. Boivin hired a lawyer to contest the demotion through the Berlin court.

When the senate met less than a month later, some members expressed their dismay at the president's decision. "We were given no documents — only a simple statement from the MPS and no statement from Boivin's side," says Ulrike Beisiegel, who was president of the University of Göttingen until 2019 and has been a member of the MPS senate since 2011. MPS regulations require that such an action is approved by the senate; a retrospective vote on Boivin's demotion was postponed until the following senate meeting, in March this year.

Ahead of the March senate meeting, Becker says that he called political representatives on the senate and advised them to request an independent investigation into the Boivin affair. This time, the senate was given

documents about the case and gave a majority vote in support of the demotion. But Beisiegel says that the documents were not discussed at the meeting. “The senate does not act like a real board, ensuring the society follows procedures,” says Beisiegel. “It is a serious problem.”

## Opaque proceedings

Boivin says that she was never given a proper hearing, and was given the details of the accusations against her only at the end of the years-long investigation. She says that the investigation did not always follow internal MPS rules and that some of the same people sit on multiple committees involved in the process.

*Nature* spoke to six Max Planck directors who had been demoted, were under investigation for non-scientific misconduct or had raised concerns about procedures internally. All had similar criticisms about the lack of transparency and perceived bias in MPS investigations, which they say involve too few independent arbiters. (All the directors asked not to be named for fear of retaliation.)

Some countries, including Denmark and Sweden, have established national agencies to investigate allegations of misconduct in researchers, to avoid issues of bias and non-transparency. Beisiegel helped to create Germany’s first national guidelines on good scientific practice and handling of misconduct in 1997, which all universities must adopt. The guidelines work well in general, she says, and scientists in Germany are attached to the freedom afforded to them by the constitution. “So I think in Germany it would be very difficult to come up with an external body,” says Beisiegel.

*Nature* **606**, 632-633 (2022)

doi: <https://doi.org/10.1038/d41586-022-01600-7>

---

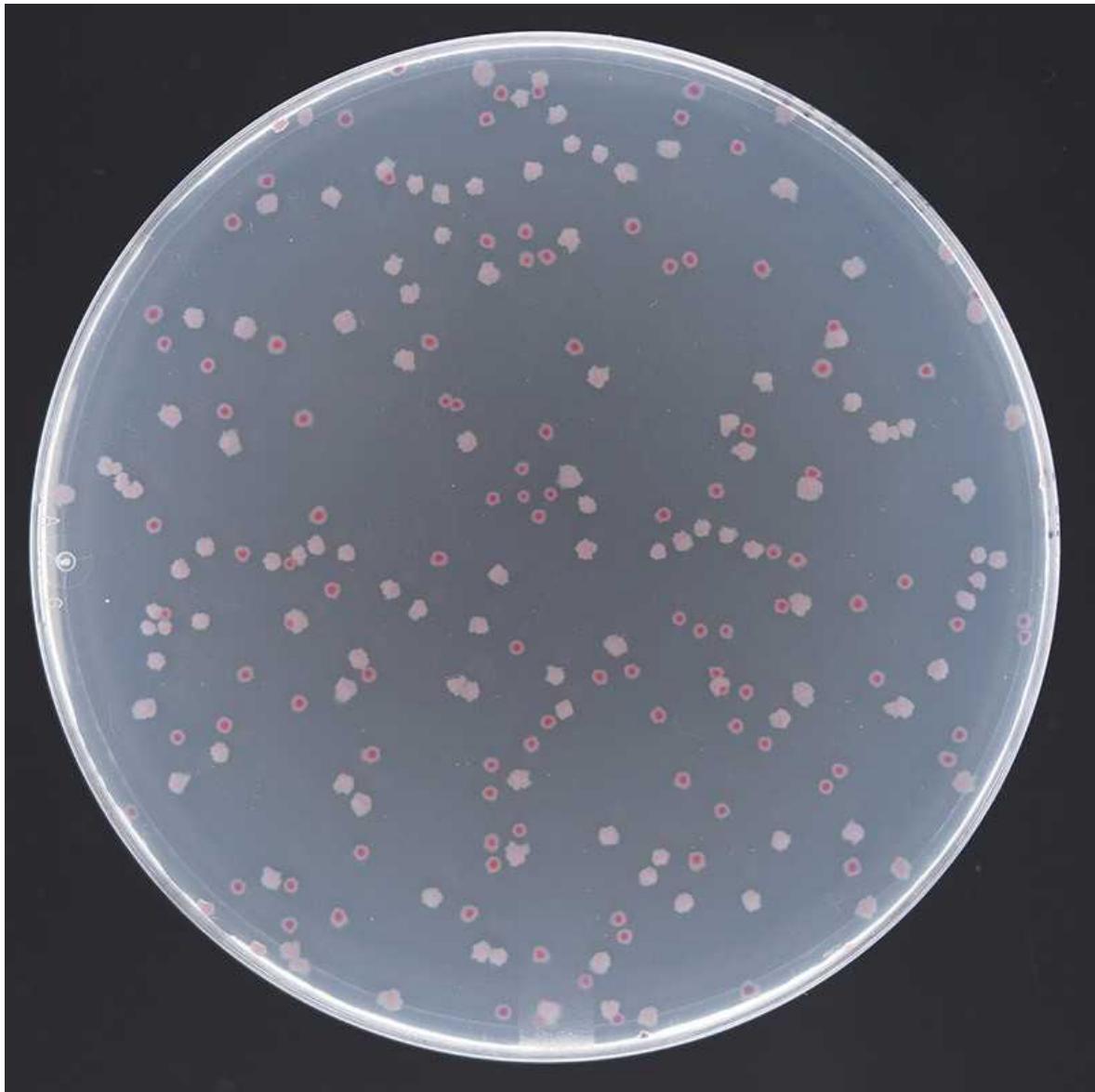
This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-022-01600-7>

- NEWS Q&A
- 14 June 2022

# Legendary bacterial evolution experiment enters new era

A laboratory has been growing 12 populations of *E. coli* since 1988 — this year, the cultures will get a new custodian.

- [Ewen Callaway](#)



*Escherichia coli* cells used in the long-term evolution experiment on an agar plate. Credit: Jeffrey Barrick

On 24 February 1988, evolutionary biologist Richard Lenski filled 12 flasks with sugary growth medium and seeded each with *Escherichia coli* bacteria. For the past 34 years, Lenski, at Michigan State University in East Lansing, and his colleagues have nurtured the bacterial cultures, refreshing growth media daily and freezing samples for future study every couple of months.

The [long-term evolution experiment \(LTEE\)](#) has become a cornerstone in evolutionary biology that researchers continue to mine for insights. During

their 75,000 generations of growth, the bacteria have made huge gains in their fitness — how fast they grow relative to other bacteria — and evolved some surprising traits.

Last month, Lenski and his lab tended to the LTEE for the last time. The 12 *E. coli* lines are now frozen in cryoprotective media and will soon be revived to begin a new life in the laboratory of Jeffrey Barrick, an evolutionary biologist at the University of Texas at Austin who first worked on the experiment in the 2000s as a postdoc in Lenski's lab.

The two spoke to *Nature* about the past, present and future of one of biology's longest-running experiments.



Jeffrey Barrick (left) will take over the running of the *E. coli* experiment from mentor Richard Lenski, who started the experiment in 1988. Credit: Laura Gerson/UT College of Natural Sciences, Sipa US/Alamy

## **What inspired you to begin the experiment?**

Richard Lenski: I like big, open-ended questions. I wanted a very simple, long experiment to look at how repeatable evolution was. The original goal was 2,000 generations. And I figured that deserves the moniker ‘the long-term evolution experiment’. I had no idea that it would go as long as it has actually gone — and hopefully will go much longer.

## **Why did you keep it going — and going and going?**

Lenski: It’s actually a very easy experiment to maintain. The amount of work for an individual on a typical day is maybe half an hour. It is 365 days a year, at least in principle, but the amount of work per day isn’t huge.

And then, of course, the bacteria were doing very interesting things over time. New technologies emerged, like the ability to sequence genomes cheaply. People like Jeff joined the lab and brought new ideas and questions. And the bacteria kept doing interesting things. I realized that it should just go on for as long as humanly possible.

## **How many generations had passed when Jeff started working on the experiment?**

Lenski: Did you join the lab in about 2007? It was probably around 40,000 generations.

Jeffrey Barrick: Sounds about right. I knew less going in about the history of the experiment — maybe less than a lot of Rich’s postdocs and graduate students. But I was at a time in my career where I had been studying evolution at the molecular level, looking at all kinds of bacterial genomes. I wanted to study evolution of whole organisms in the lab and be able to watch evolution. It’s just been something I’ve been fascinated by ever since I was an undergraduate.

## **Was it ever a challenge to keep the experiment going?**

Barrick: Like Rich said, it's fairly simple. There are some major snowstorms and other things that go on, and other infrastructure anomalies that can make it sometimes challenging, but you can always go back to the freezer, which is one really nice thing about the experiment and makes it much more doable than other experiments. People have tried long-term experiments with mice and flies and other organisms where it's really difficult if something goes wrong.

Lenski: One of the advantages of the long-term experiment is everything is so simple. We work with a chemically defined medium, we can freeze strains and revive them so that, knock on wood, I think it should continue quite smoothly in the new home.

## **What has the experiment taught us about the repeatability of evolution?**

Lenski: My bias going into the experiment was that all the strains would go off in very different directions. I was thinking that the roles of chance and contingency in evolution would have been larger than they were. And over the years, we've actually seen just striking amounts of reproducibility. So although a typical line has improved its relative fitness compared with the ancestor by maybe 70% or 80%, the variance in competitive fitness between most lines is more like just a few per cent. So they've all tremendously increased, but very similarly to one another.

But then, over the years, we've also seen quite striking divergences between the lines. Thirty thousand generations into the experiment, one of the 12 lines evolved the ability to consume citrate, instead of just glucose. And that garnered quite a lot of attention, and even some, shall we say, hostility from some people who are sceptical of the power of evolution. And after 75,000 generations , it's still the only one of the 12 lines that has evolved that ability.

**Are there big questions about evolution that you hope to answer by going for longer?**

Barrick: For a lot of bacteria that ended up in simple, constant environments — especially simple endosymbionts that live inside the cells of insects — their genomes gradually shrink over time. And I'd say one of the more surprising things to me is that these *E. coli* have been in a very constant environment, yet their genomes have not shrunk very much.

Lenski: I think part of the issue on the genome shrinkage is that that's a slow process. Thirty years and 75,000 generations — it's a drop in the evolutionary bucket. So I would guess that if we could come back — in one million years or whatever, the bacteria probably would have extremely reduced genomes. That's a reason to keep it going.



The *E. coli* has now propagated some 75,000 generations. Credit: Brian Baer and Neerja Hajela via Wikimedia Commons (CC BY-SA 1.0)

**Why did you decide to pass the torch on the LTEE?**

Lenski: I'm not going to be around forever. I think it's better to do those things now, to plan them carefully and thoughtfully. So it just made sense. I'm 65 years old, and although I don't plan to retire any time in the next few years at least, the lab is getting smaller. And one of the important things for keeping the long-term lines going is this daily rhythm. I think a lab that has half a dozen or more people is ideally suited to the weekend and holiday coverage that the experiment benefits from.

So I asked Jeff, maybe in 2018 or 2019. I have a National Science Foundation grant to run the experiment, and Jeff is now a co-principal investigator on that. And when we wrote the most recent renewal, we laid out this plan for the transfer and that it would take place around now.

## **Why did Jeff take it on?**

Barrick: I'm a big proponent of open science. This is a great resource that I want to support and share and continue. It's become kind of a common touchstone for a lot of stories about bacterial evolution. And something that people can take in so many directions. I'm excited about supporting the community.

## **Are you anxious about taking responsibility for such a long-running experiment?**

Barrick: As Rich said, it's not super-difficult to keep the experiment going. What becomes difficult is organizing all the strains in the freezer, making sure that you can send them to people, and all the paperwork involved with those sorts of things.

I don't want to have the experience where the experiment comes to my lab and then gets contaminated, and I set it two or three years back. I want to make sure that I continue things really well. But a lot of the pressure is off, because it is frozen, not only in Rich's lab and my lab, but labs in France and other places. So that takes the pressure off that I could cause an irretrievable problem. So I'm more excited about educating my colleagues about the experiment.

## Rich, what's your advice for Jeff?

Lenski: Keep calm and carry on. Frustrating things will happen. But the experiment is quite forgiving. As long as my brain is working, I'll be really excited to see what new spin-off experiments he and collaborators generate, what new analyses he and the broader community generate to make sense of what's going on with the long-term lines. Probably the most important thing Jeff will have to think about in 20 or 30 years is, who's next?

Barrick: Your job is not done, Richard. You're still the best science communicator and person for bringing the long-term evolution experiment to people and building that community. That's actually the most intimidating thing that nobody could replace at the current time. Doing the experiments is very easy.

*Nature* **606**, 634-635 (2022)

*doi:* <https://doi.org/10.1038/d41586-022-01620-3>

---

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-022-01620-3>

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

- NEWS
- 15 June 2022
- Clarification [16 June 2022](#)
- Correction [16 June 2022](#)

# Ancient DNA traces origin of Black Death

Genomes show that plague-causing bacteria found in Kyrgyzstan graves are direct ancestors of those that triggered the medieval pandemic.

- [Ewen Callaway](#)



An engraved tombstone of a person who died from the Black Death plague, from the Kara-Djigach cemetery in what is now Kyrgyzstan. Credit: P.-G.

Borbone, M. A. Spyrou *et al.*/Nature

A Silk Road stopover might have been the epicentre of one of humanity's most destructive pandemics.

People who died in a fourteenth-century outbreak in what is now Kyrgyzstan were killed by strains of the plague-causing bacterium *Yersinia pestis* that gave rise to the pathogens responsible several years later for the Black Death, shows a study of ancient genomes.

"It is like finding the place where all the strains come together, like with coronavirus where we have Alpha, Delta, Omicron all coming from this strain in Wuhan," says Johannes Krause, a palaeogeneticist at the Max Planck Institute for Evolutionary Anthropology in Leipzig, Germany, who co-led the study, published on 15 June in *Nature*<sup>1</sup>.

Between 1346 and 1353, the [Black Death laid waste to western Eurasia](#), killing up to 60% of the populace in some places. Historical records suggest that the bubonic plague emerged from the east: Caffa, on the Crimean peninsula, experienced one of the earliest-recorded outbreaks of plague during a 1346 siege by the army of the Mongol Empire. The Caucasus and other locales in Central Asia have been put forward as potential epicentres.

China hosts some of the world's greatest genetic diversity of modern *Y. pestis* strains, hinting at an East Asian origin for the Black Death. "There were all kinds of hypotheses in the literature. And it was not really known where it exactly came from," says Krause.

## Signs of the plague

Several years ago, Philip Slavin, an economic and environmental historian at the University of Stirling, UK, and a co-lead author of the study, came across records from a pair of fourteenth-century cemeteries in Kyrgyzstan that, he thought, might hold clues to the origins of the Black Death. The cemeteries, known as Kara-Djigach and Burana, held an unusually high number of tombstones dated to 1338 and 1339, ten of which made explicit reference to a pestilence.

“When you have one or two years with excess mortality, it means something funny is going on there,” Slavin said at a press briefing.

To determine whether the burials held any relevance to the later Black Death, Slavin worked with Krause to track down the remains from the Kyrgyz cemetery — which had been excavated in the 1880s and 1890s and moved to St Petersburg, Russia. The team, led by archaeogeneticist Maria Spyrou at the University of Tübingen, Germany, sequenced ancient DNA from seven people whose remains were recovered, discovering *Y. pestis* DNA in three burials from Kara-Djigach.



The Tian Shan mountain range, where researchers have found modern strains of plague-causing bacteria related to the ancient ones from Kyrgyzstan. Credit: Lyazzat Musralina

A pair of full *Y. pestis* genomes gleaned from the data showed that the bacteria were direct ancestors of strains linked to the Black Death, including a *Y. pestis* sample from a person who died in London that Krause’s team sequenced in 2011. The Kara-Djigach strain was also an ancestor of the vast majority of *Y. pestis* lineages around today — a sign, Krause says, of an

explosion in *Y. pestis* diversity shortly before the Black Death. “It was like a big bang of plague,” he said at the press briefing.

Other evidence puts the origins of the Black Death in this part of Central Asia. Among modern strains of *Y. pestis* bacteria, those sampled from marmots and other rodents in Kyrgyzstan, Kazakhstan and Xinjiang in northwest China, surrounding the Tian Shan mountain range, were most closely related to the Kara-Djigach strain. “We can’t really say it’s that village or that valley, but it’s likely that region,” says Krause.

Rodents are the natural reservoir for *Y. pestis*, and humans develop bubonic plague only when a vector such as a flea passes on the infection. Krause suspects that humans’ close contact with infected marmots sparked the Kyrgyzstan epidemic, whereas immunologically naive rat populations in Europe fuelled the Black Death.

Tian Shan makes sense as an epicentre for the Black Death, says Slavin. The region is on the ancient Silk Road trade route, and the Kyrgyzstan graves were found to contain pearls from the Indian Ocean, corals from the Mediterranean and foreign coins, suggesting that faraway goods passed through the area. “We can hypothesize that trade, both long distance and regional, must have played an important role in spreading the pathogen westward,” Slavin said.

## Medieval ‘death certificates’

Obtaining genomes from plague bacteria ancestral to those behind the Black Death is “a tremendous breakthrough”, says Monica Green, a medieval historian and independent scholar in Phoenix, Arizona. “The headstones are as close as we will ever come to ‘death certificates’. So we know this lineage of *Y. pestis* was in existence then.” But she’s less sure of the study’s conclusion that the plague’s ‘big bang’ occurred around the time of the Kyrgyzstan deaths in 1338–39. Green has hypothesized, on the basis of genetic, ecological and historical evidence, that the thirteenth-century expansion of the Mongol Empire catalysed the spread and diversification of *Y. pestis* strains responsible for the later Black Death.

Sharon Dewitte, a bioarchaeologist at the University of South Carolina in Columbia, says the work opens the doors to studying the Black Death — and the wider outbreak it was part of, known as the second plague pandemic — beyond Europe. She's keen to compare demographic and mortality patterns from people who died of the plague in Kara-Djigach with those from European Black Death cemeteries.

“Having more plague samples from ancient Asia and China will be super interesting in terms of adding even more evidence to the Asian origin of the first and second [plague] pandemics,” adds Simon Rasmussen, a computational biologist at the University of Copenhagen who has analysed ancient *Y. pestis* sequences.

Krause hopes to analyse remains from China to see how a pandemic that so scarred Europe reverberated in East Asia, he says. “We would really like to get the Eastern part of the story.”

*Nature* **606**, 635-636 (2022)

doi: <https://doi.org/10.1038/d41586-022-01673-4>

## Updates & Corrections

- **Clarification 16 June 2022:** This story has been updated to make clear that Monica Green used ecological and historical evidence, as well as genetic evidence, to reach her conclusions about the spread of the Black Death.
- **Correction 16 June 2022:** This story originally quoted Johannes Krause saying “it was really known where it exactly came from”; that should have said “it was not really known where it exactly came from”. The text has been updated.

## References

1. Spyrou, M. A. *et al. Nature* <https://doi.org/10.1038/s41586-022-04800-3> (2022).

---

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-022-01673-4>

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

- NEWS
- 10 June 2022

# African researchers lead campaign for equity in global collaborations

Cape Town statement on research partnerships between the global north and south will highlight unethical practices and offer advice to scientists.

- [Holly Else](#)



South Africa's former public protector Thuli Madonsela delivered a keynote speech at the seventh World Conference on Research Integrity.Credit: University of Cape Town

Researchers at the seventh World Conference on Research Integrity, in Cape Town, South Africa, have been hammering out the equity issues plaguing science partnerships that span the global north–south divide.

Several sessions at the event were dedicated to the design of a soon-to-be-published document called the Cape Town Statement on fostering research integrity. The conference — the first to be held in Africa — ran from 29 May to 1 June.

The statement will offer guidance on how researchers from low- and middle-income countries can become equal partners in international projects. The organizers hope that having a set of principles for fair and equitable partnerships will help scientists from the global south to speak out against unfair practices. These include not being properly credited or pursuing research questions imposed by collaborators from the global north that do not benefit local communities.

Although not all international collaborations are problematic, unfair and inequitable practices are rife, says Lyn Horn, who heads the office of research integrity at the University of Cape Town. She is formulating the statement together with a group of researchers and ethicists. “Even people and funders with very good intentions perhaps don’t understand how entrenched some of their practices and processes are,” she says.

Amos Laar, a public-health researcher at the University of Ghana in Accra, says that the development of the statement on African soil by African researchers will empower scientists to challenge inequity.

## Unequal partnerships

Researchers in low- and middle-income countries often collaborate with peers from wealthier ones. This can bring advantages in the form of secure funding, which might not be available locally, and better career prospects as a result of working with prestigious institutions. Their partners from higher-income countries benefit, too, because they gain access to local communities and expertise.

But these collaborations can be fraught with inequity. Sometimes the research is funded and led by overseas scientists who arrive with a fully formed research question that either doesn't address the issues of local people, or disregards their customs and traditions. Local researchers often have little involvement outside data collection and fieldwork, and in some cases they are not given fair credit for their contribution. An analysis of 94 papers on COVID-19 published in 10 of the leading medical journals that contained content relating to Africa<sup>1</sup> found that two-thirds of the authors were not from the continent, and that one in five articles had no African authors. Researchers might also be unable to access the resulting data, if they are uploaded to servers in institutions the global north.

In the worst cases, local researchers have no involvement at all — a phenomenon called [helicopter science](#) — or do work that would not be ethically approved in wealthier countries, a situation known as ethics dumping.

“The statement is going to open up space for people to proactively engage on these issues,” says Francis Kombe, director of EthixPERT, an organization in Pretoria, South Africa, that promotes ethical research, and one of those involved in the formulation of the Cape Town statement.

On 16 May, a paper outlining the issues and possible solutions<sup>2</sup> was published on the OSF preprint server for discussion at the conference. It includes a preliminary list of key values and principles — such as accountability, mutual respect and fairness — that are important for fostering equality and could be incorporated in the statement. The document also proposes actions that funders, research-team leaders, institutions, journals, publishers and scientists involved in these partnerships can take to redress the balance.

Those leading the discussions at the conference will now finesse the text of the statement, and hope to publish the work in a journal later this year.

“The fact that we have contributed to the generation of these principles makes us feel that they are ours, we can lead by them,” says Laar. “Otherwise it is pushed down upon us from elsewhere.”

## Ethical guidelines

The statement comes amid a wider reckoning about the power dynamics in international research collaborations. Some wealthier institutions have launched initiatives to try and make their partnerships with researchers in low- and middle-income countries more equitable.

Researchers from a network of Finnish institutions that do international-development research are working on a set of ethical guidelines to help to decolonize academic partnerships.

And the London School of Hygiene & Tropical Medicine — a prestigious UK research university — is currently reviewing its global partnerships. [Last year, an independent review commissioned by the university found evidence of racism and inequality there](#). Now, its researchers are working on a formal policy about how best to engage with partners in less-wealthy nations, says Patricia Henley, head of research governance and integrity at the institution.

“We’re taking a step back and really looking at what we are doing, why are we doing it and how can we do it better,” she says. The policy will include ways to make sure that research questions stem from the local community.

Institutions still face hurdles when trying to work more equitably, she points out. Many research funders in the United Kingdom still insist that scientists from the wealthy nations lead the work and approve the ethics, for example. “Why does it have to be so focused on the northern institution? I think that’s a throwback from the colonial days,” Henley says.

Back in Cape Town, the pressure is now on, Horn says. The final version of the statement should not be “too broad and too wishy washy”, and should emphasize how unfair practices can affect the integrity, validity and trustworthiness of research, she adds.

But change will not happen overnight, and the first steps will be small ones, says Kombe. “The statement cannot solve everything. Once the statement is out, it’s going to plant a seed in everyone’s mind that is going to germinate

into something that is going to have a very long lasting impact on the research enterprise as a whole.”

*Nature* **606**, 636 (2022)

*doi:* <https://doi.org/10.1038/d41586-022-01604-3>

## References

1. Naidoo, A. V., Hodkinson, P., Lai King, L. & Wallis, L. A. *BMJ Glob. Health* **6**, e004612 (2021).
2. Horn, L. *et al.* Preprint at OSF Preprints  
<https://doi.org/10.31219/osf.io/bf286> (2022).

---

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-022-01604-3>

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

- NEWS
- 13 June 2022

# New COVID drugs face delays as trials grow more difficult

Fewer people are eligible for the massive studies needed to test treatments for severe COVID-19.

- [Saima May Sidik](#)



The antiviral tablet Paxlovid sharply reduces risk of hospitalization and death from COVID-19. Credit: Chris Sweda/Chicago Tribune/Tribune News Service via Getty

After [two years of breakneck research](#), scientists have amassed a collection of therapies to treat people with COVID-19. But now, researchers fear that development of new treatments could falter as the clinical trials needed to test them become increasingly difficult.

Vaccinations in many places have led to a decline in severe disease, shrinking the pool of potential study participants. Hesitance to enrol in trials is rising, and the existence of potent treatments is making statistical analysis more difficult, too.

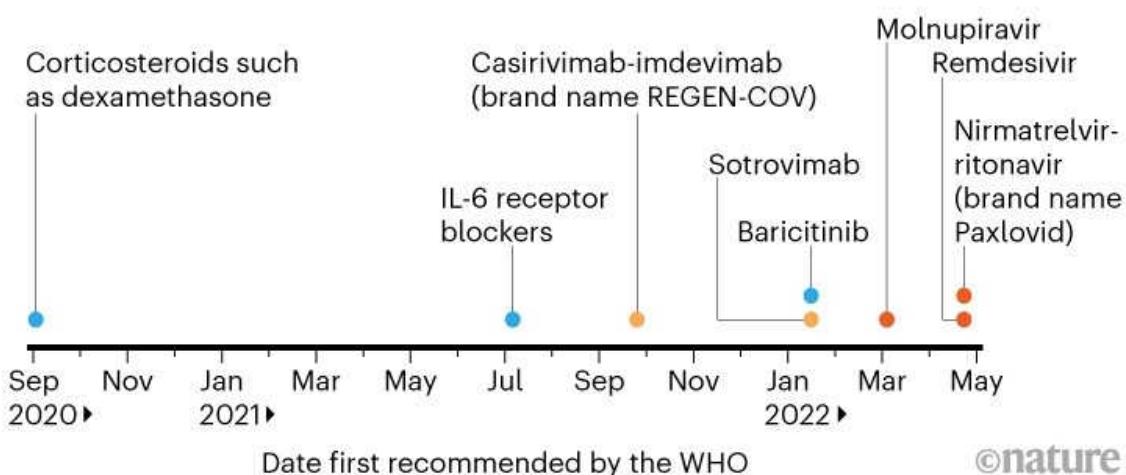
“It was definitely easier to do research in the past. Now you’ve got to design a study that meets the standards of care doctors want to do, and patients want to do. And it’s a lot harder,” says Elizabeth Hohmann, an infectious-disease expert at Massachusetts General Hospital in Boston.

Doctors treating people with COVID-19 can choose from roughly half a dozen types of therapy that have been recommended by the World Health Organization, or by national authorities such as the US Food and Drug Administration (see ‘Virus-taming tools’). Among them are steroids, synthetic antibodies and antiviral tablets. Some cut the risk of death for those already in hospital. Others lower the odds of having to be hospitalized at all. Death rates are dropping in some countries that are fortunate enough to have access to these treatments, and modelling<sup>1</sup> suggests that widespread antiviral treatment could prevent the majority of COVID-19 deaths.

## VIRUS-TAMING TOOLS

The virus that causes COVID-19 was identified only in early 2020, but the World Health Organization has already recommended more than half a dozen treatments for the disease. Still other therapies have been recommended by domestic agencies such as the US National Institutes of Health.

● Antiviral   ● Monoclonal antibody   ● Anti-inflammatory



Source: World Health Organization, US National Institutes of Health

But, in many areas, the available therapies are limited in supply and high in cost. There's also the looming spectre of resistance to drugs such as the antiviral Paxlovid (nirmatrelvir–ritonavir), developed by Pfizer in New York City. Researchers worry that progress in establishing new treatments will stall, even as many parts of the world are left without treatment options.

## Shrinking pool

Thanks largely to vaccines, certain hard-hit countries have seen death rates drop precipitously. In Brazil, for example, where deaths were once running at 3,000 deaths a day, the figure has declined to less than 200 a day. Yet that welcome news can complicate trials.

At the beginning of the pandemic, health researcher Edward Mills at McMaster University in Hamilton, Canada, and his colleagues set up a trial in Brazil to learn whether existing drugs could prevent the most serious outcomes of COVID-19. When they launched the trial, called TOGETHER,

in early 2020, the share of study participants who eventually died or had to be hospitalized was 16%. But the number dropped to 3–5% after vaccines became available. Before they could continue testing whether certain drugs prevent severe outcomes, the organizers therefore had to enrol more people who were in danger of becoming critically ill. That meant expanding the trial to further sites — in South Africa, Pakistan, the Democratic Republic of the Congo and Rwanda.

## Another type of hesitancy

Scientists also worry that even those people who do qualify for trials are more reluctant to take part than they would have been at the beginning of the pandemic.

When Hohmann began overseeing a trial called ACTT to test COVID-19 treatments in early 2020, recruitment was quick: ill people had no better option. By April 2020, the trial had enrolled 1,062 people. And by the end of 2020, it had shown that the antiviral drug remdesivir speeds recovery and prevents death<sup>2</sup>.

But Hohmann says that, as effective treatments such as remdesivir became available, it became more and more difficult to recruit participants for subsequent trials. Many people feel safer sticking with the established regimen, which today includes both remdesivir and the steroid dexamethasone, than trying an experimental drug as well.

“It just takes a much more adventurous person to step onto that third drug,” Hohmann says. It also takes some civic-mindedness to sign up for a trial if your life isn’t at risk, Hohmann says, and she suspects that the stress and upheaval of the pandemic has whittled away at potential participants’ altruism.

Hesitance of another sort might have affected a Canadian clinical trial for the drug losartan as a treatment for severe COVID-19. The majority of people in Canada became vaccinated during 2021, so most of the people available to join the losartan trial were unvaccinated. That could explain the rise in the share of people who were invited to join the study but declined:

18% in mid-2021 and 35% by the end of the year<sup>3</sup>. Trial organizers suspect that the same factors that [make people hesitant to receive vaccines](#), such as distrust of mainstream medicine, also make them disinclined to take experimental drugs.

## Statistical complexity

As treatments have multiplied, so too has the complexity of the statistical calculations needed to determine whether a new drug is effective. As a result, researchers might need to recruit more trial participants, which takes more time.

That's precisely what has happened for the organizers of the ongoing PRINCIPLE trial, which tests whether repurposed drugs can speed recovery or keep infected UK residents out of the hospital. All study participants receive the current standard of care, meaning that doctors are free to prescribe treatments in addition to the drug being tested. That dilutes any difference in outcome between participants taking the placebo and those taking the treatment under study, says Ly-Mee Yu, a medical statistician at the University of Oxford, UK, and PRINCIPLE's lead statistician. Smaller differences mean that researchers need to work with larger groups of participants, and therefore trials take longer.

Hohmann notes that if researchers want to compare a new, highly effective drug against Paxlovid, the powerful antiviral that is now the leading treatment for early COVID-19, they will need to recruit a huge number of trial participants to discern a statistically significant difference between the two treatments. "You'd have to have a real game changer to take on Paxlovid," she says.

But researchers might need to embrace the difficulties of finding a Paxlovid challenger. Recent experiments on cells<sup>4,5</sup> suggest that Paxlovid-resistant strains of the virus might arise — a stark reminder that no matter how complex the playing field becomes, the virus is defining the rules of the game.

*doi:* <https://doi.org/10.1038/d41586-022-01602-5>

## References

1. Matrajt, L., Brown, E. R., Cohen, M. S., Dimitrov, D. & Janes, H. Preprint at medRxiv <https://doi.org/10.1101/2021.11.10.21266139> (2022).
2. Beigel, J. H. *et al.* *N. Engl. J. Med.* **383**, 1813–1826 (2020).
3. Russell, J. A., Walley, K. R., Kalil, A. C. & Fowler, R. *Ann. Amer. Thor. Soc.* <https://doi.org/10.1513/AnnalsATS.202203-250VP> (2022).
4. Zhou, Y. *et al.* Preprint at bioRxiv <https://doi.org/10.1101/2022.06.06.494921> (2022).
5. Jochmans, D. *et al.* Preprint at bioRxiv <https://doi.org/10.1101/2022.06.07.495116> (2022).

---

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-022-01602-5>

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

# Opinion

- **Bernard Bigot (1950–2022)** [16 Jun 2022]
  - Director-general of ITER, the world's largest fusion experiment.
- **Sustainable small-scale fisheries can help people and the planet** [20 Jun 2022]
  - Artisanal fishing can improve livelihoods, boost nutrition and strengthen food systems, but fishers' input is needed at local, national and global levels.

- OBITUARY
- 16 June 2022

# Bernard Bigot (1950–2022)

Director-general of ITER, the world's largest fusion experiment.

- [Eisuke Tada](#)



Credit: ITER Organisation

Bernard Bigot said, “We try to understand nature, and then we try, within human limitations, to recreate what nature shows us.” He was referring to

fusion energy, to which he devoted the final years of his life. The statement illuminates equally well his contributions to society as a chemist, physicist, educator, administrator and inspirational leader. He was a committed civil servant, adviser to French presidents, former administrator-general of the French Alternative Energies and Atomic Energy Commission (CEA) and, from 2015, director-general of the ITER Organization in southern France, which oversees the world's largest experimental fusion reactor. He has died in office aged 72.

For Bigot, the universal availability of massive carbon-free energy sources was a central goal in his mission to bring about a better world. ITER, a 35-country collaboration to build the first industrial-scale fusion facility, clearly fell into this vision. When he arrived, the project was experiencing difficulties that threatened its very existence. Starting with a series of structural and managerial reforms, including a move to centralized decision making and the creation of cross-organizational project teams, Bigot put it back on track. He eventually conquered the twin challenges of the machine's complexity and its manufacture being dispersed across three continents. Many have argued that he saved ITER and restored momentum to the quest to make fusion power part of society's future.

Born in Blois in the Loire Valley, France, in 1950, Bigot won admission to the École Normale Supérieure (ENS) of Saint-Cloud, where he pursued two teaching degrees. In 1979, he graduated with a PhD in theoretical chemistry from the Pierre and Marie Curie University in Paris. He led research teams at the ENS and at the Institute of Research on Catalysis in Lyon, in both cases building on his growing reputation as a modeller of real systems, such as heterogeneous catalysis systems, as well as photochemistry and condensed-state chemistry. He organized the transfer and transformation of the ENS from Saint-Cloud to Lyon in the mid 1980s and was appointed director of the century-old institution in 2000.

Yet his sense of a duty to advance the common good led him from academic research to civil service, where he made his greatest contributions. Both the fission of atomic nuclei and their fusion are potential sources of almost unlimited energy. Fission is now globally industrialized: in France, 56 reactors account for 70% of electrical power generation. Fusion is still in its experimental phase, and ITER is designed to demonstrate its technological

and industrial feasibility. At its heart is a tokamak, a doughnut-shaped vacuum chamber roughly the weight of the Eiffel Tower, surrounded by magnetic coils. In the chamber, a gas of hydrogen isotopes at high temperature forms a plasma in which nuclei can fuse in the same reaction that powers the Sun, producing energy as heat.

Bigot's contributions to carbon-free energy began in fission, as he helped to shape France's nuclear policy. From 2002 to 2003, he was chief of staff to French research minister (and former astronaut) Claudie Haigneré. He was high commissioner for atomic energy from 2003 to 2009, and then headed the CEA until 2015. During his tenure, the commission officially added 'Alternative Energies' to its name and became a driving force in the ITER project.

Bigot's international network, and his natural diplomatic talent, proved precious when he came to steer ITER — a collaboration between China, Europe, India, Japan, South Korea, Russia and the United States — through sometimes treacherous waters. His greatest challenge came in 2020, with the onset of COVID-19. ITER's massive components — superconducting magnets the weight of a Boeing 747 aeroplane, precision-forged and welded pieces of steel the size of Stonehenge, pieces of the vacuum vessel that took more than five years to make — were beginning to arrive for integrated assembly. Bigot rallied ITER's partners to a consensus that the project would continue without pause, a decision that ultimately proved successful.

Bigot and I were involved in ITER from its very early days, well before the ITER Organization was born. We sat on different sides of the table, he for France and I for Japan, discussing where we would build the facility. It was my joy that I was able to work with him, sitting at the same table, after he called me to join him in 2015. We talked for 15 minutes in his office at 7.45 every morning, before the day became hectic. During these calm, quiet, yet efficient meetings, I learnt a lot about his way of living, thinking and working, together with his incredible integrity, passion and energy.

The French call men of Bigot's nature and attitude *moine-soldat*, meaning that they combine in their personality and behaviour the simplicity (sometimes perceived as austerity) of the monk and the determination, loyalty and dedication of the soldier. Bernard Bigot was a monk-soldier, but

also a man of a deep human warmth and sensitivity as a husband, father and friend. To his grandchildren, he explained his role at ITER as like that of a builder of a medieval cathedral who would not live to see the final masterpiece, but whose work created hope that would drive others to complete the task.

An educator at heart, Bigot had a rare talent for explaining the most complex ideas. Government executives, media representatives and others will remember how fusion seemed simple and thrilling when he explained it.

As committed to the present as he was engaged in the future, Bigot remained deeply attached to a number of educational institutions and foundations, including the Maison de la Chimie Foundation — a Paris-based centre for chemistry — of which he was president. Preparing younger generations for the challenges they will face was always central to his preoccupations. Well beyond ITER and the international research community, his premature passing will be felt as a tremendous loss.

*Nature* **606**, 645 (2022)

*doi:* <https://doi.org/10.1038/d41586-022-01681-4>

---

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-022-01681-4>

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

- COMMENT
- 20 June 2022

# Sustainable small-scale fisheries can help people and the planet

Artisanal fishing can improve livelihoods, boost nutrition and strengthen food systems, but fishers' input is needed at local, national and global levels.

- [Sheryl L. Hendriks](#)



Women in the fishing village of Ngor Dakar in Senegal unload part of the catch from a traditional fishing boat. Credit: Nic Bothma/EPA/Shutterstock

More than three billion people rely on the ocean to make a living, most of whom are in developing countries. For some 17% of the world's population, fisheries and aquaculture provide the main source of animal protein. For the least-developed countries, fish contributes about 29% of animal protein intake; in other developing countries, it accounts for 19%<sup>1</sup>.

As the global population increases, the demand for seafood is expected to rise, too. Already, Africa and Asia have seen fish production double over the past few decades. Globally, fish consumption is set to rise by around 15% by 2030<sup>2</sup>.

Although ocean ecosystems are strained by climate change, overfishing and more, studies nevertheless suggest that seafood can be expanded sustainably to meet future food demands<sup>3</sup>. Last year, international efforts promoting this approach included the [Blue Food Assessment](#) (a joint initiative of 25 research institutions) and the United Nations Food Systems Summit.

Success will depend on small-scale fisheries. Small operations tend to deliver both food and income directly to the people who need them most, and locals have a strong incentive to make their practices sustainable. What's more, these fisheries can be remarkably efficient. Almost everything that hand-to-mouth fisheries catch is consumed. By contrast, around 20% of the fish caught by industrial fleets is estimated to be wasted, mainly because of unwanted by-catch<sup>4</sup>. So, whereas large-scale operators land more fish, small-scale fisheries provide a larger share of the fish that is actually consumed.

Small fishers rarely have the right resources to expand their operations, or even to survive. If they do scale up, they might lose some of their current advantages or engage in the same harmful practices as do large commercial fisheries. Managed with care, however, small fisheries could provide win-wins for livelihoods and the environment. Making this happen should be high on the agenda at the [UN Ocean Conference in Lisbon this month](#).

As someone who has studied food security and policymaking for decades, here I suggest ways to support and strengthen artisanal fishing operations.

## Small reforms

The potential and importance of small-scale fisheries has been increasingly recognized over the past decade. In 2014, the UN Food and Agriculture Organization (FAO) provided voluntary guidelines to support sustainable small-scale fisheries, aimed at improving food security and eradicating poverty. A forthcoming report by the FAO, Duke University in Durham, North Carolina, and the non-profit organization WorldFish, headquartered in Penang, Malaysia, will conclude a remarkable initiative to collate case studies, questionnaire results and data sets to help get fishers a seat at policymakers' tables. The UN General Assembly has declared 2022 the International Year of Artisanal Fisheries and Aquaculture.

Most nations already have management policies for marine ecosystems that provide for small-scale fisheries. In India, Indonesia, Malaysia and Sri Lanka, for example, there is a ban on trawling within about 8 kilometres of the coastline to prevent industrial fishers from scooping up large catches, which protects those regions for local fishers. Countries such as Costa Rica ease access by exempting small-scale fisheries from licences, and Angola exempts subsistence and artisanal fishers from paying licensing fees<sup>5</sup>.

But this is not enough. Small-scale fishers' rights to access are often poorly defined, ineffectively enforced or unfairly distributed<sup>4</sup>. The boundaries of exclusive economic zones (EEZs) — the parts of the coast belonging to a given nation — are often poorly policed, and large-scale vessels regularly swoop in and take sea life through bottom trawling, something that small fishers seldom practice. Large-scale bottom-trawlers account for 26% of the global fisheries catch, with more than 99% of that occurring in the EEZs of coastal countries<sup>6</sup>. Even when there are well-meaning policies to protect local fishers, foreign vessels can take advantage. For instance, a 2018 investigation by the Environmental Justice Foundation in London found that around 90% of Ghana's industrial fishing fleet was linked to Chinese ownership, despite Ghanaian laws expressly forbidding foreign ownership or control of its boats. Clearer definitions of the terms fisher, fishing and fishing vessel to make provisions for small-scale operators could help, in part, to avoid such abuse.

Government subsidies also require reform. One estimate found that large-scale fishers receive about three-and-a-half times more subsidies than small-scale fishers do<sup>7</sup>. This widens the existing advantages of large operations in terms of vessels and gear, infrastructure (including cold storage), processing capacity and access to cheap fuel. By giving large-scale fishers the capacity to catch even more, it can have the perverse effect of encouraging overfishing<sup>8</sup>. Instead, subsidies and other funds should be directed towards small-scale fishers to let them expand their access to markets, while keeping them from adopting the negative practices of large-scale operations.

## More for consumption

The total global loss and waste from fisheries is estimated at between 30% and 35% annually<sup>1</sup>. This could increase as smaller operations broaden their markets. A 2015 estimate of the Volta Basin coast in West Africa attributed 65% of fish-production losses to a lack of technology and good manufacturing practices, and to a lack of infrastructure such as decent roads and cold storage<sup>9</sup>. The study found that fish were rarely lost to physical damage during the process; most waste resulted from spoilage. Such losses limit the sale of fish locally and to distant markets.

Public and private investment in cold-storage facilities and processing equipment (such as for drying, fermentation, pickling or smoking) could help. Current funding for fishery conservation projects comes from development partners, regional banks, the World Bank, private foundations and other agencies — with some entities also providing microloans to small-scale fisheries — but these efforts are uncoordinated and inadequate.

One promising strategy is to pair international or national funding with direct contracts for feeding programmes linked to schools, hospitals and similar facilities. Such arrangements would provide small fisheries with large, consistent markets and storage infrastructure that boosts local consumption and does not incentivize overfishing.



Artisanal fishers at a fish-processing cooperative in Santa Rosa de Salinas, Ecuador. Credit: Camilo Pareja/AFP/Getty

Other strategies pair local fishers with conservation efforts. As fishing operations scale up, fish entrails and other waste cannot simply be thrown into the sea: care must be taken not to contaminate the environment. One option is to fund ecosystem-restoration projects that also benefit local fisheries. For example, the *Mikoko Pamo ja* (Mangroves Together) project in Gazi Bay, Kenya, restores and conserves degraded mangrove forests, which act as nurseries for young fish. The restoration thus earns saleable carbon credits while enhancing nearby fishery grounds for the local community.

Consumers could support small fisheries by buying local, because shorter supply chains mean more income for the fishers. The use of ecolabels — which seek to promote sustainably managed fisheries by certifying that a product has a reduced environmental impact — could also encourage consumer adoption, and help consumers to make informed choices.

However, such certification is costly to obtain and maintain, and requires compliance, monitoring and reporting. Certification can distort market

opportunities, effectively excluding small enterprises from entering international markets. These programmes can also have unintended consequences: most certification programmes focus on environmental sustainability and pay less attention to social responsibility elements, such as fairness in access to resources, markets and wages.

Instead, simple incentive programmes could be implemented by funders, managers and local governments trying to promote sustainable fisheries. For example, local markets could display a rating system for individual fishers or small entrepreneurs. This could include various elements of sustainability other than environmental ones — such as providing information on the type of fishing gear, location of the catch and freshness. Promoting the rating as a social responsibility concept would inform consumers of the need to support sustainable fisheries. The rating system could be conducted by community members trained in inspection and enforcement.

## Local control

Diverse efforts are needed to protect small fisheries' access and to boost local consumption and reduce waste, and must be tailored to local community conditions. The 2021 UN Food Systems Summit was a 'people's summit' that elevated roles for Indigenous peoples and civil-society groups, yet the voice of fishing communities was notably absent.

Few governments take an integrated approach to the development, implementation and enforcement of policies. For example, policies governing urban development tend not to consider the implications on the ocean, fish and fishers. In the late 2000s, for instance, fishers were initially denied access to traditional public fishing zones along the beach front in Durban, South Africa, following upgrades to the port and the development of a private marina and hotel. (Fishers later reclaimed some of the zones after protests and engagement with the authorities<sup>10</sup>.)

Cooperatives can help on several fronts: by coordinating fishing activities, sharing information (about weather, sea conditions or fish movement) and advocating effectively for human and social rights. For instance, CoopeSoliDar, a small-scale fisheries management cooperative in San José,

Costa Rica, has helped to strengthen collective action to sustainably use molluscs, alleviate poverty and strengthen the representation of women and young people in community decision-making. Governments can help by creating a legal framework to establish cooperatives and include them in decisions to manage marine resources.

Local communities can also stand up for themselves. For example, a class action by a group of 5,000 artisanal fishers in South Africa in 2004 argued against a policy they said did not give them recognition or access to food and fishing rights that were established in the country's constitution. The court ruled in the group's favour in 2007, and the resulting legal framework granted small-scale fishers collective community fishing rights, recognizing community members as bona fide fishers<sup>[11](#)</sup>.

## Integrated inputs

Small fisheries do not operate in isolation. Unlike terrestrial resources, the ocean is an extensive, global commons without clear territorial boundaries. Issues as diverse as climate change, ocean acidification, overfishing and pollution by nutrients and plastics and other chemicals all affect local fishers. But such system interactions get scant attention when fisheries policies focus on a single seafood stock or individual fishing area.

Whereas the concept of integrated land management has been part of the development agenda for a few decades, integrated marine management is only now emerging. To work, it must involve all relevant stakeholders, including small-scale fishers.

A context-specific strategy in the Seychelles is a leading example of such integration. Communities, financing partners and the government worked together to create the Seychelles Marine Spatial Plan Initiative, which protects 30% of the archipelago's waters and boosts climate resilience. The Seychelles faces significant threats from rising sea levels and warmer air and water temperatures that put fisheries, infrastructure, tourism and its rich biodiversity at risk.

In an example in the Coral Triangle region (encompassing Indonesia, Malaysia, Papua New Guinea, the Philippines, the Solomon Islands and East Timor), local communities gave their input to a marine protection plan. This led to a greater understanding of how practices such as overfishing and taking undersized stock sustains marine and coastal resources, and how managing these helps to address food security, climate change and threats to marine biodiversity. Such cooperation between fishing communities and governments in managing marine protected areas is essential to the preservation of future fish stocks (see [go.nature.com/3xvkqxj](https://go.nature.com/3xvkqxj)).

Fishers should be actively engaged in relevant meetings held by the UN and national and local councils, so that they can weigh in on matters that affect fishing access, their livelihoods and environmental concerns. Both fishers and organizers must help to build empowerment mechanisms to make sure their voices are heard, such as providing translation services and scheduling meetings at accessible locations. This is important not just for the fishers' human rights, but also because much can be learnt from artisanal fishers' local knowledge.

Moves that would, for instance, restrict the fishing season or areas so that stocks or biodiversity can recover should include compensation mechanisms that will secure fishers' cooperation and livelihoods. Social-protection measures such as food and income assistance can also help to tide fishers over.

When fish swim in schools, they move more efficiently, forage better and are protected from predators. The same might be said for small-scale fishers, but those networks should extend to local and international communities, too. Collaborative problem-solving and an integrated food system can deliver seafood protein, sustainably, to a world that increasingly needs it.

*Nature* **606**, 650-652 (2022)

doi: <https://doi.org/10.1038/d41586-022-01683-2>

## References

1. Food and Agriculture Organization of the United Nations. *The State of World Fisheries and Aquaculture 2020: Sustainability in Action* (FAO, 2020).
2. Organisation for Economic Co-operation and Development & the Food and Agriculture Organization of the United Nations. *OECD-FAO Agricultural Outlook 2021–2030* (OECD-FAO, 2021).
3. Costello, C. *et al.* *Nature* **588**, 95–100 (2020).
4. High-level Panel of Experts on Food Security and Nutrition. *Sustainable Fisheries and Aquaculture for Food Security and Nutrition* (HLPE, 2014).
5. Nakamura, J., Chuenpagdee, R. & El Halimi, M. *Mar. Policy* **129**, 104568 (2021).
6. Steadman, D. *et al.* *New Perspectives on an Old Fishing Practice: Scale, Context and Impacts of Bottom Trawling* (Blue Ventures, 2021).
7. Schuhbauer, A., Skerritt, D. J., Ebrahim, N., Le Manach, F. & Sumaila, U. R. *Front. Mar. Sci.* **7**, 539214 (2020).
8. Hopewell, K. & Margulis, M. E. *Nature Food* **3**, 2–3 (2022).
9. Diei-Ouadi, Y. *et al.* *Strengthening the Performance of Post-Harvest Systems and Regional Trade in Small-Scale Fisheries: Case Study of Post-Harvest Loss Reduction in the Volta Basin Riparian Countries*. FAO Fisheries and Aquaculture Circular No. 1105 (FAO, 2015).
10. Maharaj, B. *Local Econ.* **32**, 744–762 (2017).
11. Nthane, T. T., Saunders, F., Gallardo Fernández, G. L. & Raemaekers, S. *Sustainability* **12**, 743 (2020).

# Research

- [\*\*Solar flare accelerates nearly all electrons in a large coronal volume\*\*](#) [ 08 Jun 2022]
  - By evolving spatially resolved distributions of thermal and non-thermal electrons in a solar flare in a large coronal volume, it is shown that nearly all electrons experienced a prominent acceleration.
- [\*\*Observation of a correlated free four-neutron system\*\*](#) [ 22 Jun 2022]
  - Experiment based on knocking out an alpha particle from a high-energy helium isotope shows a resonance-like structure that is consistent with a quasi-bound tetraneutron state existing for a very short time.
- [\*\*Continuous Bose–Einstein condensation\*\*](#) [ 08 Jun 2022]
  - Continuous, indefinitely lasting Bose–Einstein condensation, sustained by amplification through Bose-stimulated gain of atoms from a thermal bath, creates a continuous-wave condensate of strontium atoms.
- [\*\*Many-body theory of positron binding to polyatomic molecules\*\*](#) [ 22 Jun 2022]
  - A many-body theory of binding interactions between positrons and polar and nonpolar molecules is developed, showing agreement with experimental data up to within 1%.
- [\*\*Engineering topological states in atom-based semiconductor quantum dots\*\*](#) [ 22 Jun 2022]
  - Precision-engineered devices consisting of a linear array of ten quantum dots are used to realize both the trivial and topological phases of the many-body Su–Schrieffer–Heeger model.
- [\*\*Organic bipolar transistors\*\*](#) [ 22 Jun 2022]
  - An organic bipolar junction transistor composed of highly crystalline rubrene thin films has a device architecture that could be used in organic electronics with greatly improved high-frequency performance
- [\*\*The source of the Black Death in fourteenth-century central Eurasia\*\*](#) [ 15 Jun 2022]
  -
- [\*\*The role of NSP6 in the biogenesis of the SARS-CoV-2 replication organelle\*\*](#) [ 12 May 2022]

- The non-structural protein NSP6 in SARS-CoV-2 has a key role in viral replication by zippering the endoplasmic reticulum membrane to establish connectors between the double-membrane vesicles of the viral replication organelle and the endoplasmic reticulum.
- **Cohesin-mediated loop anchors confine the locations of human replication origins** [ 08 Jun 2022]
  - A study shows that the three-dimensional conformation of the human genome influences the positioning of DNA replication initiation zones, highlighting cohesin-mediated loop anchors as essential determinants of their precise location.

---

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

- Article
- Open Access
- [Published: 08 June 2022](#)

# Solar flare accelerates nearly all electrons in a large coronal volume

- [Gregory D. Fleishman](#) ,
- [Gelu M. Nita](#),
- [Bin Chen](#),
- [Sijie Yu](#) &
- [Dale E. Gary](#)

*Nature* volume **606**, pages 674–677 (2022)

- 3526 Accesses
- 393 Altmetric
- [Metrics details](#)

## Abstract

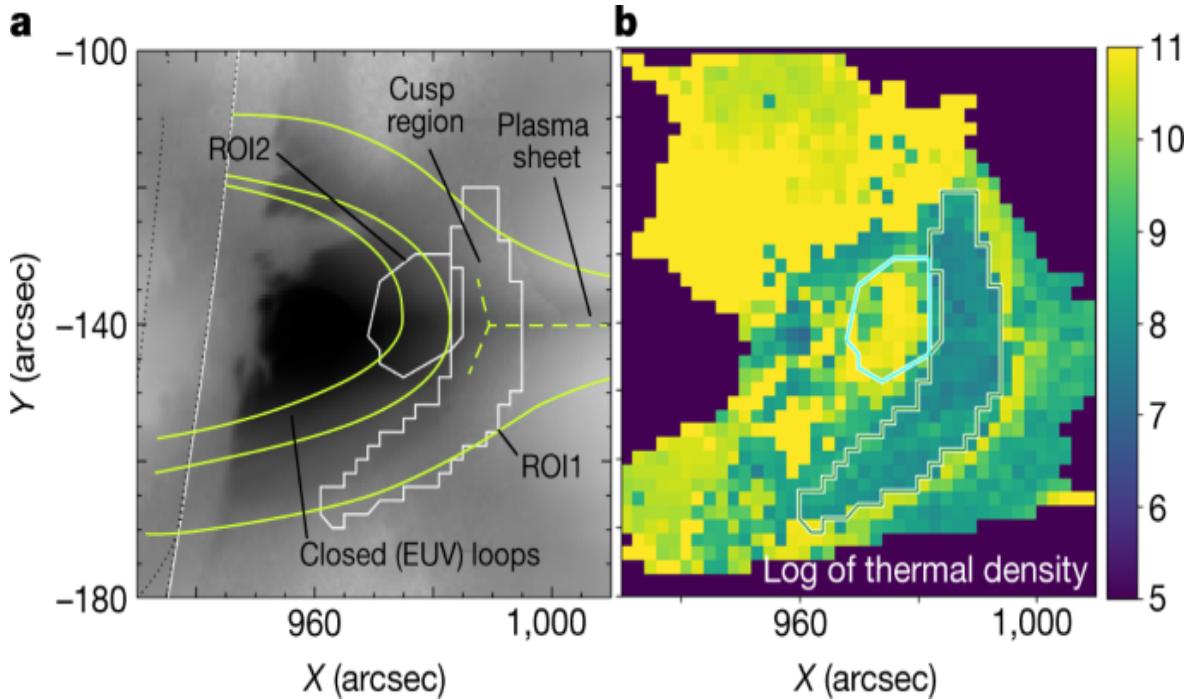
Solar flares, driven by prompt release of free magnetic energy in the solar corona<sup>1,2</sup>, are known to accelerate a substantial portion (ten per cent or more)<sup>3,4</sup> of available electrons to high energies. Hard X-rays, produced by high-energy electrons accelerated in the flare<sup>5</sup>, require a high ambient density for their detection. This restricts the observed volume to denser regions that do not necessarily sample the entire volume of accelerated electrons<sup>6</sup>. Here we report evolving spatially resolved distributions of thermal and non-thermal electrons in a solar flare derived from microwave observations that show the true extent of the acceleration region. These distributions show a volume filled with only (or almost only) non-thermal

electrons while being depleted of the thermal plasma, implying that all electrons have experienced a prominent acceleration there. This volume is isolated from a surrounding, more typical flare plasma of mainly thermal particles with a smaller proportion of non-thermal electrons. This highly efficient acceleration happens in the same volume in which the free magnetic energy is being released<sup>2</sup>.

## Main

The microwave analysis is performed using imaging spectroscopy data from the Expanded Owens Valley Solar Array (EOVSA) described in detail elsewhere<sup>2</sup>. We use independent pixel-to-pixel and time-to-time spectral model fitting of these microwave imaging spectroscopy data to obtain evolving, spatially resolved distributions (maps) of suprathermal and thermal electrons. These maps pinpoint the location and shape of the evolving acceleration region in a large solar flare that occurred on 10 September 2017 (Fig. 1). This now famous flare has attracted extraordinary attention as it demonstrates several ‘textbook’ flare properties, which were observed with unprecedented coverage and resolution across the electromagnetic spectrum<sup>7,8,9,10,11,12,13,14</sup>.

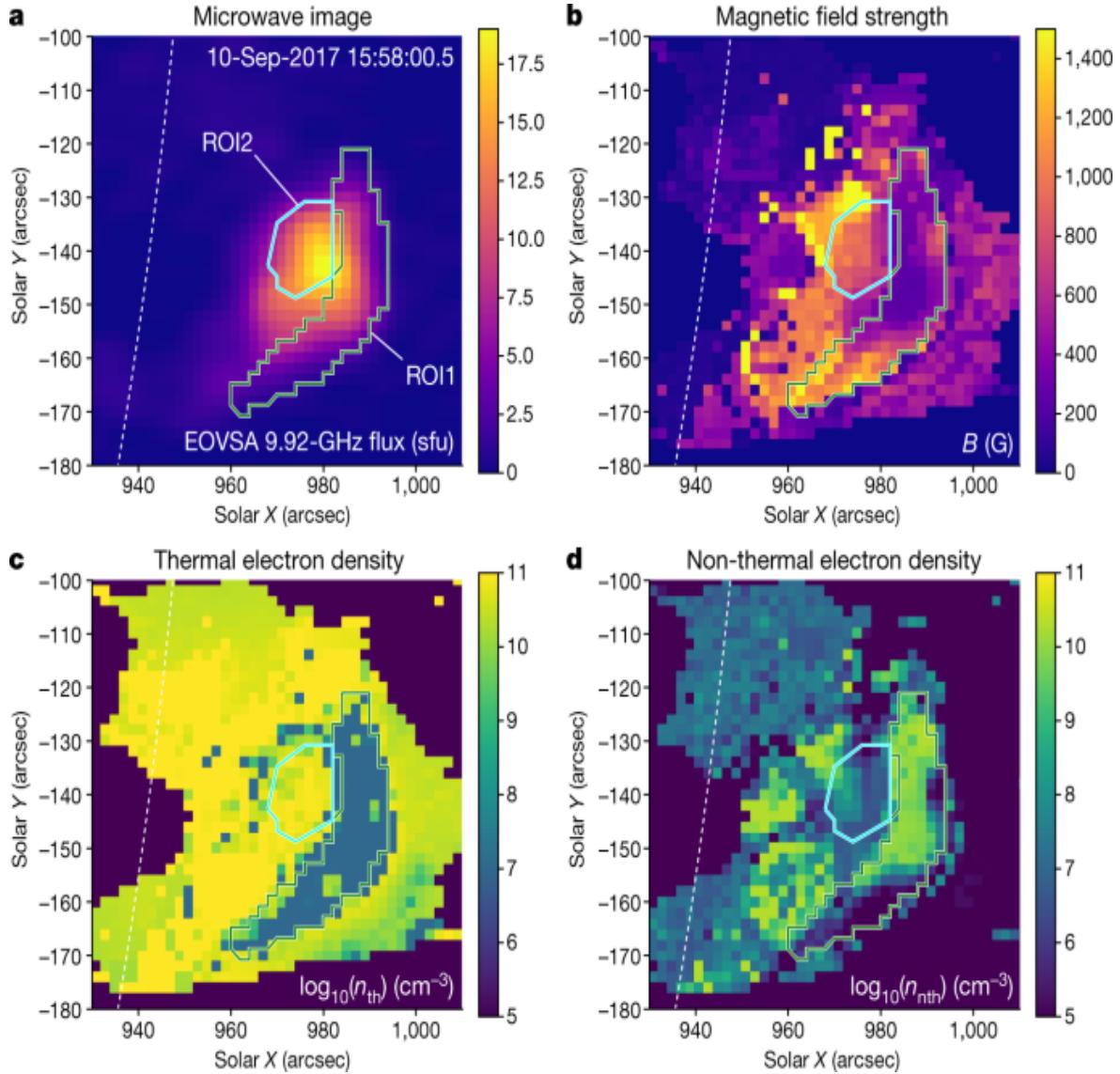
**Fig. 1: Thermal plasma in the context of magnetic morphology in the 10 September 2017 solar flare.**



**a**, The plot shows EUV brightness distribution (image) in the context of the hand-drawn magnetic field lines outlining closed, post-reconnection loops, cusp region including a so-called Y-point and the vertical plasma sheet. ROI1 and ROI2 used in the analysis are also shown. **b**, Distribution of the median values of the thermal plasma density, log scale, obtained from the microwave data using MCMC simulations; see [Methods](#). Strong depletion of the thermal number density is apparent inside ROI1.

Figure 2 shows an example of these parameter maps for one time frame; the entire evolution is illustrated in Supplementary Video S1. Figure 2 also shows two regions of interest (ROIs), ROI1 and ROI2, kept fixed for all analysed time frames, which inscribe two areas having the most reliable spatially resolved spectra and, thus, the most reliable model spectral fitting diagnostics (see [Methods](#)). ROI1 inscribes the area in which the fast and strong release of coronal magnetic energy has been measured<sup>2</sup>, whereas ROI2 is a reference area of more typical flare plasma, outside the acceleration region, to be used for comparison.

**Fig. 2: Spatial distributions of flare parameters.**



**a**, EOVSA map at 9.92 GHz taken at 15:58 UT. **b–d**, Maps of the magnetic field (**b**), thermal plasma density (**c**) and non-thermal plasma density (**d**) derived for the same time from the bulk model spectral fitting (see [Methods](#)). ROI1 inscribes the hole in the distribution of thermal plasma, which also corresponds to a peak in the number density of the suprathermal electrons. ROI2 inscribes a reference area; see [Methods](#). The dotted arc shows the solar limb. Note that, owing to different processing, panel **c** differs slightly from Fig. [1b](#), which was produced by the more thorough but time-consuming MCMC method; see [Methods](#). sfu, solar flux unit.

We focus on ROI1, in which the fast, strong release of magnetic energy occurred during the main flare phase<sup>2</sup>, thus pinpointing the exact energy

release region. Figure 2d shows that ROI1 inscribes an extended area (corresponding to an estimated volume of about  $1.67 \times 10^{27} \text{ cm}^3$ ; see [Methods](#)), in which the number density of suprathermal electrons with high energies above 20 keV is very large—up to around  $10^{10} \text{ cm}^{-3}$ . By contrast, the number density of thermal electrons in ROI1, shown in Fig. 2c, is undetectably small (see [Methods](#)): the map contains an extended thermal density ‘hole’ roughly coinciding with ROI1. This directly implies that the number density of suprathermal electrons is much larger than that of the thermal electrons in the region in which the release of magnetic energy takes place.

Supplementary Video [S1](#) demonstrates that the gap in the thermal electron distribution holds for the entire duration of the analysed four-minute episode around the peak of the flare, although its shape evolves and shows an overall outward motion (to the right in the figure), and it continues to match the region of enhanced suprathermal electron density. These spatio-temporal evolutions show that, during the entire episode, ROI1 and ROI2 differ fundamentally in character: suprathermal electrons dominate in ROI1, whereas the thermal electrons dominate in ROI2. The suprathermal electrons in ROI1 seem to have been accelerated in place, rather than transported there from elsewhere (see [Methods](#)). Therefore, ROI1 combines three properties: (1) fast release of a large amount of magnetic energy<sup>2</sup>; (2) depletion of thermal plasma; and (3) presence of a dense population of suprathermal electrons, presumably accelerated owing to the magnetic energy release. This combination of properties implies that we have resolved the heart of the solar flare—the exact acceleration region that places strong constraints on the physical mechanism driving the acceleration of electrons in the flare. Indeed, any mechanism capable of producing a suprathermal particle population has to extract a fraction of charged particles from the thermal plasma pool and increase individual energies of those particles greatly. As a result, at this acceleration stage, the number of accelerated suprathermal particles increases at the expense of the thermal particles, whose number density proportionally decreases. In our case, ROI1 has a lower (possibly much lower) than 10% proportion of thermal plasma in a region of high suprathermal electron density (see [Methods](#)). This means that a large fraction, essentially all, of the thermal electrons originally present in this volume has been converted to the suprathermal electron population during

(and, presumably, owing to) this energy release. We conclude that the magnetic energy release in the solar flare offers a highly efficient engine for particle acceleration, which is capable of converting essentially all ambient electrons with thermal energies (for example, less than  $\approx 1$  keV) into a suprathermal population of electrons with high energies exceeding 20 keV (see [Methods](#)).

The release of free magnetic energy, quantified by the fast decay of the magnetic field at the rate  $\dot{B} \approx 5 \text{ G s}^{-1}$ , has been suggested to be driven by turbulent magnetic reconnection within an extended volume of the cusp region of the flare<sup>2</sup>. This is motivated by the inferred highly enhanced turbulent magnetic diffusivity,  $v \approx 10^{15} \text{ cm}^2 \text{ s}^{-1}$ , and the associated strong electric field,  $E \approx 20 \text{ V cm}^{-1}$ , in that extended volume<sup>2</sup>.

The fundamental force capable of producing work on charged particles is the electric force. The acceleration efficiency is specified by a balance between the energy gain due to the electric field and the energy loss through collisions, which defines a critical value of the electric field, called the Dreicer field,  $E_D$  (ref. [15](#)). The condition for runaway acceleration is  $E \gg E_D$ , which is called a ‘super-Dreicer’ electric field. The electric field inferred from the magnetic field decay<sup>2</sup>, about  $20 \text{ V cm}^{-1}$ , is many orders of magnitude larger than the estimated Dreicer field, which is  $E_D \approx 10^{-4} \text{ V cm}^{-1}$ .

To support the simultaneous acceleration of literally all ambient electrons in a macroscopic volume such as ROI1, this strong super-Dreicer field must be present over a substantial portion of ROI1. As noted above, this is consistent with the observed simultaneous decay of the magnetic field over the entire ROI1 (ref. [2](#)), indicative of turbulent magnetic reconnection, in which the dissipation of the magnetic energy takes place throughout the volume. This is in contrast to the alternative view that all acceleration takes place in one or a few isolated points (X-points or O-points<sup>16</sup>) favourable for macroscopic reconnection.

From this work, there emerges a consistent picture of particle acceleration in the magnetic energy release region: (1) the decay of the magnetic field owing to turbulent magnetic reconnection produces a strong super-Dreicer

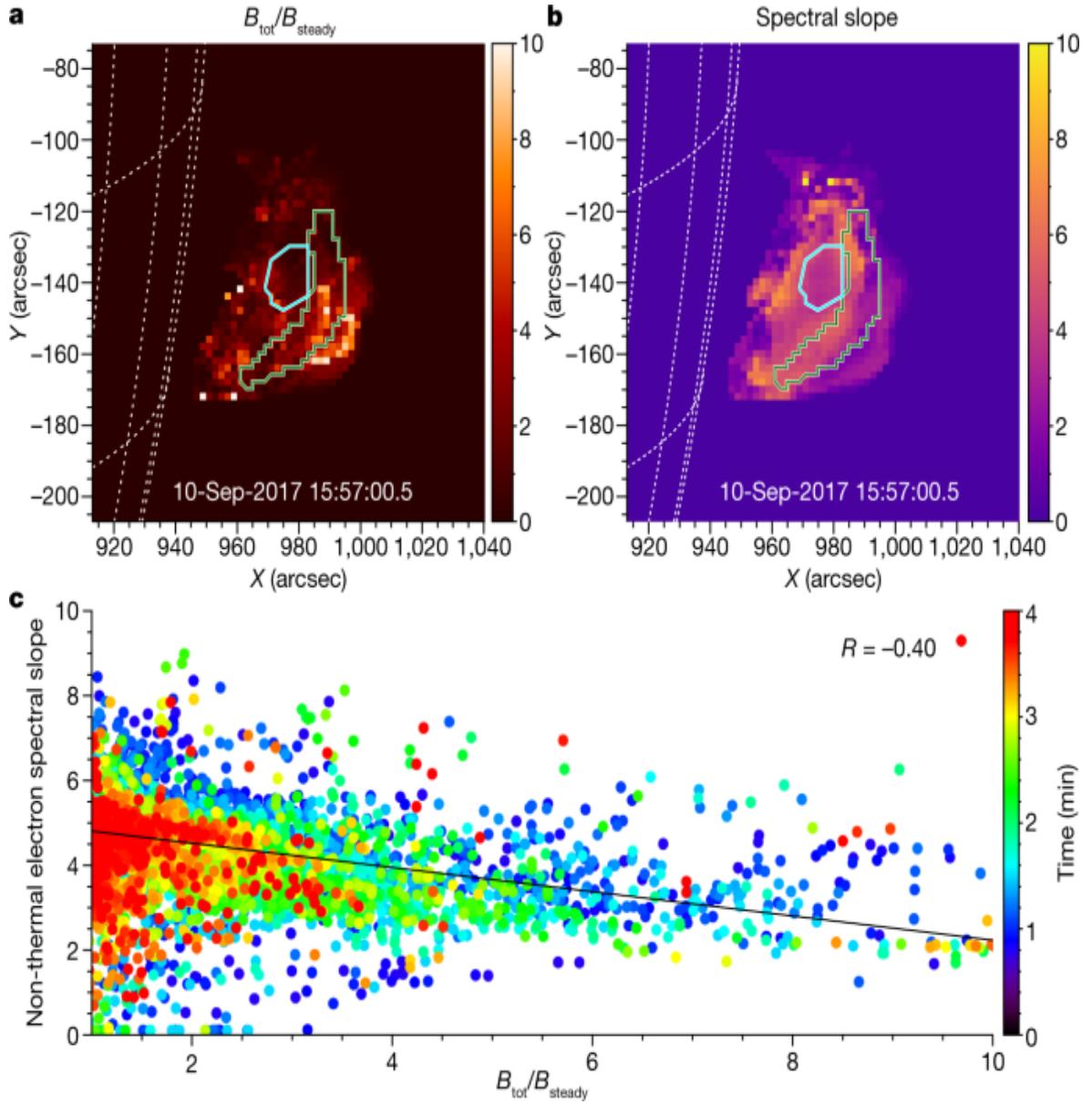
electric field over an extended volume; (2) this strong electric field works over literally all ambient particles, which boosts their energies up to 20 keV and higher; (3) this acceleration process is so efficient and persistent that it does not leave any measurable thermal plasma component compared with the highly dominant suprathermal component.

Models of particle acceleration owing to magnetic reconnection, including 2D and 3D particle-in-cell simulations, as well as a new large-scale kinetic simulation approach kglobal<sup>17,18</sup>, suggest that efficiency of the acceleration is linked to a ratio of the reconnecting (dissipating) component  $B_{\text{rec}}$  of the total magnetic field and the remaining component, called the guide field,  $B_g$  (ref. <sup>19</sup>), does not explicitly take part in the reconnection. According to the models<sup>17,18,20,21,22</sup>, efficient acceleration requires that  $B_{\text{rec}}$  is larger than  $B_g$ . We check this expectation with our data. Although we cannot properly separate these two components observationally, we can compare the observed total magnetic field,  $B_{\text{tot}}(t)$ , at a given time and location with its value  $B_{\text{steady}}$  near the end of the decay period, when it becomes steady<sup>23</sup>. We estimate  $B_{\text{steady}}$  at each location in ROI1 as the mean  $B$  evaluated over the last 20 s of the four-minute episode<sup>2</sup>.  $B_{\text{steady}}$  serves as an estimate for the magnetic field component that does not participate in the energy release process, which includes  $B_g$ . Then, for each pixel, we form a ratio  $r_B(t) = B_{\text{tot}}(t)/B_{\text{steady}}$ , in which  $B_{\text{tot}}(t)$  is the instantaneous value of the magnetic field inferred from the spectral model fit, which includes both decaying and steady components of the magnetic field. If the observed  $\langle \{r\}_{\{B\}} \gg 1 \rangle$ , then it is probable that  $\langle \{B\}_{\{\text{rec}\}}/\{B\}_{\{g\}} \gg 1 \rangle$  as well (small guide field case); thus,  $\langle \{r\}_{\{B\}} \gg 1 \rangle$  could be viewed as a good proxy for efficient acceleration. We focus on the first two minutes of our four-minute episode, in which the condition  $\langle \{r\}_{\{B\}} \gg 1 \rangle$  holds for many pixels and times.

We investigate the relationship between parameters  $n_{\text{nth}}$  and  $\delta$  of the suprathermal electron component derived from the spectral fit and the ratio  $r_B$ , in which the non-thermal number density exceeds the thermal number density. We did not find any correlation of the number density of the suprathermal electrons with  $r_B$ . This is consistent with the observed strong

efficiency of the acceleration, which results in virtually all ambient electrons being accelerated. What is correlated with acceleration efficiency is the power law spectral index  $\delta$ , as shown in Fig. 3. The correlation is such that a larger  $r_B$  (proxy for the small guide field case) implies a smaller spectral index (harder energy spectrum), thus validating the theoretical expectations<sup>18</sup>. A simplistic interpretation of this relationship is that having more free magnetic energy (larger  $r_B$ ) permits acceleration to higher energies, thus producing a flatter distribution of the accelerated electrons over energy.

**Fig. 3: Spectral slope of the suprathermal component versus reconnecting magnetic field.**



**a, b**, Maps of the inferred  $r_B = B_{\text{tot}}/B_{\text{steady}}$  ratio (**a**) and the suprathermal electron energy spectral slope (**b**), averaged over the first five time frames (20 s) of the full four-minute interval. The ROI1 and ROI2 regions are marked, respectively, by green and cyan contours in both panels. **c**, Dot symbols show the spectral index  $\delta$  obtained from the ROI1 model fit for each pixel and time versus the running value of the  $r_B$  ratio. The correlation data are plotted for the analysed four-minute interval, colour coded as time increases, as shown by the colour scale. The solid line shows the linear fit to the data over the first two minutes, when most of the energy release takes place, which corresponds to a correlation coefficient  $R = -0.40$ . An animated

version of this figure (Supplementary Video [S2](#)) showing all time frames used in this analysis is provided as supplemental material.

In addition, our observations show that the suprathermal electrons, generated in a region in which virtually all ambient electrons are accelerated, remain almost perfectly isolated from the surrounding cooler plasma for a time period much longer than the source transit time, even though the system does not contain any ‘solid walls’ that would hold suprathermal particles in. This means that the system contains a highly efficient physical process or magnetic topology that traps the suprathermal particles within the volume they occupy. Otherwise, the suprathermal particles would become much more uniformly mixed with the ambient thermal particles, which is not observed. An important process capable of providing this trapping is enhanced angular diffusion that reduces the particle mean free path<sup>[24](#)</sup>. Such diffusion is due to particle scattering by the turbulent magnetic field, which is also responsible for acceleration of the particles. Although the need for this enhanced diffusion is strongly suggested by the observations, the important characteristics of the corresponding turbulent magnetic field, such as their spectral and spatial structure and evolution, remain unknown and call for dedicated modelling.

## Methods

### Overview

In this study, we used the dataset from the EOVSA<sup>[25](#)</sup> described in an earlier paper<sup>[2](#)</sup>. The model spectral fitting, its parameters and their uncertainties were described in the supplementary materials to that paper. The parameters used to create the evolving maps of the thermal and suprathermal electrons in the flare region are from the same spectral fits as those used for the magnetic field maps reported there. Here we used these maps of electron parameters to investigate the spatially resolved structure and evolution of the electron acceleration in the spatial area that showed the most prominent decay of the coronal magnetic field<sup>[2](#)</sup>.

### Spatially resolved microwave spectra and selection of ROIs

Extended Data Figure 1 shows a representative set of the observed spatially resolved microwave spectra from pixels with an area of  $2'' \times 2''$  (about  $2.1 \times 10^{16} \text{ cm}^2$ ) and associated model spectral fits distributed over the flare region. For reference, the central panel shows a single microwave image at 9.92 GHz taken at 15:58 UT, which corresponds to the main peak of the flare. (For the microwave image, the instrumental beam is  $113''.7/f(\text{GHz}) \times 53''.0/f(\text{GHz})$ . A circular restoring beam with full width at half maximum of  $87''.9/f(\text{GHz})$  was used, which is about  $9''$  for 9.92 GHz shown in the figure.) The spectral fitting uses the model of the gyrosynchrotron source function with the account of the free-free component<sup>2</sup>. We performed this model spectral fitting over all 60 time frames and over all pixels in these 60 map cubes, assuming a source depth along the line of sight (LOS) of 5.8 Mm (this corresponds to  $8''$  on disk, which is a scale of features (loops) seen in the flare images). The primary ROI, ROI1, indicated by the green contour, includes 137 image pixels that, under the same LOS depth assumptions, correspond to an estimated volume of about  $1.7 \times 10^{27} \text{ cm}^3$ . Consequently, the reference ROI, ROI2, shown by the cyan contour, which encloses 49 pixels, corresponds to an estimated volume of around  $6.0 \times 10^{26} \text{ cm}^3$ . The numbered points are pixels whose spectra and fits are shown in the other eight panels of the plot.

ROI1 inscribes the area in which the most prominent decay of the magnetic field has been detected, a small portion of which was analysed in the earlier paper<sup>2</sup>. Here we analyse the entire ROI1 as it shows a coherent depletion of the thermal plasma and a high density of suprathermal electrons. The spatially resolved spectra (for example, pixels P1 and P4) from an upper portion of ROI1 have high signal-to-noise ratio and their spectral peaks occur within the frequency range observed by the EOVSA. As a result, the model spectral fitting diagnostics using such spectra are the most robust (see the next section). In the bottom portion of ROI1, the spectra have lower signal-to-noise ratio (see example in pixel P6), especially at high frequencies, which can result in larger uncertainties of the spectral index that quantifies the suprathermal electron distribution over energy (see supplemental materials in the earlier paper<sup>2</sup>).

In the reference area ROI2, the signal-to-noise ratio is also high. The spectral peak is outside the EOVSA frequency range, indicative of high

magnetic field in ROI2. The model spectral fitting of such spectra typically yields a reliable estimate of the thermal number density, whereas the magnetic field and suprathermal electrons are recovered with larger uncertainties (see the next section).

Four other spectra from the figure corners show spectra from pixels P3, P5, P7 and P8. The signal-to-noise ratios are not large there; however, the fits are within the uncertainties and the spectra show expected trends: the spectral peak frequency is high from P3 and P8 locations close to the solar limb (which means high magnetic field strength), whereas the peak frequency is lower from higher locations P5 and P7 (which implies lower magnetic field strength). We note that because of high uncertainties of the data in the four ‘corner’ cases, the uncertainties of the derived physical parameters are also large there. Although we present parameters from all these fits in Fig. 2, we restrict our quantitative analysis to the most reliable spectra and fits from ROI1 and ROI2 and, hence, those four spectra are excluded.

## MCMC validation of the spectral model fit

The main reported result, that the number density of high-energy electrons is much larger within ROI1 than that of the thermal plasma, is based on the model spectral fitting of the microwave data. Here we use the Markov chain Monte Carlo (MCMC) simulations, implemented by an open-source Python package emcee<sup>26</sup>, to derive statistical distributions of the model fit parameters to quantify the confidence of this finding. This approach explores the full multidimensional space of the model fit parameters to both provide parameter distributions and show correlations between them. For this reason, it is much more time consuming than the speed-optimized GSFIT approach<sup>2</sup>, with which the bulk model spectral fitting has been performed. We restrict our MCMC analysis to all pixels in a single time frame, the same as shown in Fig. 2, which takes considerably longer than the GSFIT analysis of the entire 60-frame time sequence, but a comparison of the MCMC result in Fig. 1b with the bulk fitting in Fig. 2c shows that the results are comparable and fully consistent.

Thermal and suprathermal electrons affect the microwave spectrum differently. The suprathermal electrons gyrating in the ambient magnetic field are responsible for generation of the microwave emission. In the optically thin regime (high frequencies), the contributions of each individual electron add up incoherently; thus, the microwave flux level of the emission is proportional to the number density of the suprathermal component. In the optically thick regime (low frequencies), the flux of the microwave emission is determined by the energy of the electron population responsible for the emission at a given frequency. For these reasons, the microwave diagnostics of the suprathermal electrons is robust, provided that both low-frequency and high-frequency spectral ranges are available.

The thermal electrons contribute much less to the radiation intensity. Their main effect on the microwave radiation spectrum is due to dispersion of electromagnetic waves; simplistically speaking, due to the index of refraction. In the plasma, the index of refraction depends on the plasma frequency  $\omega_p$ , which is defined by the number density of the ambient free electrons:

$$\frac{\omega_p^2}{n_{\text{tot}} m} = \frac{4 \pi e^2}{m} \quad (1)$$

in which  $e$  and  $m$  are the charge and mass of the electrons and  $n_{\text{tot}}$  is the total number density of all ambient free electrons—both thermal  $n_{\text{th}}$  and suprathermal  $n_{\text{nth}}$ :

$$n_{\text{tot}} = n_{\text{th}} + n_{\text{nth}} \quad (2)$$

As  $n_{\text{th}}$ , and thus  $n_{\text{tot}}$ , increases, the microwave flux decreases at low frequencies, as illustrated in Supplementary Video [S3](#). Thus, the diagnostic of  $n_{\text{th}}$  is primarily based on the microwave spectral shape at low frequencies. If  $n_{\text{th}} \gg n_{\text{nth}}$ , then  $n_{\text{tot}} \approx n_{\text{th}}$ , offering the diagnostics of the thermal electron number density.

The MCMC analysis of a spectrum (from a pixel inside ROI2) that yields a well-constrained thermal number density is shown in Extended Data Fig. 2. The figure layout is as follows. The stand-alone upper-right panel shows a measured spectrum from a pixel within ROI2 (open circles with error bars) and a set of theoretical trial spectra (blue) consistent with the data. The panels placed over the diagonal show statistical distributions (histograms) of the trial fits for all six model parameters. The remaining panels show correlations between all possible pairs of these parameters. In this case, the distribution of thermal plasma number density is very narrow; thus, this parameter is well constrained (see also the next section). This is due to the well-measured low-frequency part of the spectrum, whose deviation from a simple power law permits this thermal density diagnostic, as explained above. By contrast, other parameters have broader statistical distributions and, thus, they are not that well constrained. This is due to the absence of the optically thin part of the measured spectrum, because the spectral peak extends beyond the EOVSA frequency range. Although the distribution of the suprathermal electron number density is broad, its relatively low most-probable value is consistent with the dominance of the thermal electrons,  $n_{\text{th}} > n_{\text{nth}}$ .

The case when  $n_{\text{th}} \ll n_{\text{nth}}$  is more problematic for the thermal plasma diagnostics, because now  $n_{\text{tot}} \approx n_{\text{nth}}$  and the thermal plasma density is defined by the difference:

$$\begin{aligned} \$\$ \{n\}_{\{\{\text{rm}\{\text{th}\}\}\}} = & \{n\}_{\{\{\text{rm}\{\text{tot}\}\}\}} - \{n\}_{\{\{\text{rm}\{\text{nth}\}\}\}} \\ & \{n\}_{\{\{\text{rm}\{\text{tot}\}\}\}}, \$\$ \end{aligned} \quad (3)$$

which is the intrinsically less constrained given uncertainties of the inputs. Thus, if the contribution of the suprathermal electrons to the total ambient density dominates, it is problematic to obtain well-constrained values of the thermal number density separately. In such a situation, we can only confidently conclude that  $n_{\text{th}} \ll n_{\text{nth}}$ , which would—in fact—confirm that most of the available ambient electrons have been accelerated to high energies. The results of the MCMC simulations for a pixel from ROI1 are shown in Extended Data Fig. 3, which has the same layout as Extended Data Fig. 2. Here the spectrum contains the peak. The distributions of the

magnetic field, suprothermal electron density and their spectral index are narrow; thus, these parameters are well constrained. The suprothermal electron number density is high, on the order of  $n_{\text{nth}} \approx 10^{10} \text{ cm}^{-3}$ . By contrast, the distribution of the thermal plasma number density is broad. It favours low  $n_{\text{th}}$  values, falling steeply for higher values. These distributions show that the thermal density contribution to the total ambient number density  $n_{\text{tot}}$  is undetectable compared with the non-thermal one, thus confirming that  $n_{\text{th}} \ll n_{\text{nth}}$ : the median values of  $n_{\text{th}}$  are less than 5–10% of  $n_{\text{nth}}$  and even the upper limit values computed as  $(\langle n \rangle_{\{\rm th\}} + 1\sigma_{\langle n \rangle_{\{\rm th\}}})$  are less than about 30% of  $n_{\text{nth}}$  at many pixels within ROI1.

The maps of the thermal and suprothermal electron densities obtained from the MCMC simulations for the entire field of view are shown in Extended Data Fig. 4. They agree within the uncertainties with those obtained using GSFIT in Fig. 2. This confirms the reliability of the results derived using the fast model spectral fitting method used in GSFIT. One apparent disagreement between Extended Data Fig. 4a and Fig. 2c is the thin line of enhanced thermal density just to the right from ROI1 in the MCMC case. Although this feature is also present in Fig. 2c, it is made less apparent because the density falls less steeply, extending the light yellow colours higher in altitude and reducing the contrast. The reason for this different appearance of the maps is that Fig. 2c shows the most probable parameter value from the GSFIT analysis, whereas Extended Data Fig. 4a shows the median value from the corresponding statistical distribution of the parameter from the MCMC simulations (compare Extended Data Fig. 3 and Extended Data Fig. 2). When the uncertainties of the derived parameters are small (their statistical distribution is narrow), then the GSFIT value is very close to the median MCMC value. However, in the area to the right of ROI1, uncertainties of the derived parameters are larger, resulting in the different appearance of these maps, even though the values are consistent with each other within uncertainties, as has been said.

Extended Data Figure 4c illustrates the dominance of the suprothermal component in ROI1 by showing  $\log(n_{\text{th,max}}/n_{\text{nth}})$ , in which  $n_{\text{th,max}}$  is represented as the median value of  $(\langle n \rangle_{\{\rm th\}} + 1\sigma_{\langle n \rangle_{\{\rm th\}}})$

$\}_{\{n\}}_{\{\{\rm rm\{th\}\}\}}\})$ ) of  $n_{\text{th}}$  in MCMC. A diverging colour map is selected for this plot, in which white colour means  $\log(n_{\text{th,max}}/n_{\text{nth}}) = 0$ . The blue/white region shows up as a distinctive feature of ROI1, with the ratio  $\log(n_{\text{th,max}}/n_{\text{nth}})$  ranging from 10% to 30%.

Note that the non-thermal number density  $n_{\text{nth}}$  is sensitive to the value of the low-energy cut-off  $E_{\min}$ , which we adopted to be fixed at 20 keV in GSFIT. In our MCMC test, we allow this parameter to vary. The assumption that  $E_{\min} = 20$  keV is proved valid in most regions of the map except in ROI1 (see the map of MCMC constrained  $E_{\min}$  in Extended Data Fig. 4d), in which the median values of  $E_{\min}$  reach 40–50 keV (see the sensitivity of the gyrosynchrotron spectrum to  $E_{\min}$  in Supplementary Video S3). Although such a concentration of non-thermal electrons can be owing to either acceleration in place or confinement of a transported electron population from elsewhere (for example, the X-point above)<sup>11</sup>, the map of  $E_{\min}$  shows that it is about two times larger in ROI1 than in the surroundings, which is rather difficult to account for without bulk electron acceleration in ROI1. The simultaneous decay of magnetic field in this same region is further support for this. We thus conclude that the suprathermal electrons in ROI1 not only have a higher number density  $n_{\text{nth}}$  but are also accelerated in bulk to a higher energy well separated from the thermal, Maxwellian component. In general, having larger  $E_{\min}$  may imply smaller  $n_{\text{nth}}$  for the same spectral slope. However, the cross-correlation plots between the parameters shown in the bottom row of Extended Data Fig. 2 demonstrate that  $E_{\min}$  correlates with  $\delta$  in such a way that larger  $E_{\min}$  corresponds to larger  $\delta$  (softer spectra). As a result of this correlation,  $n_{\text{nth}}$  does not correlate with  $E_{\min}$ ; thus, the conclusion of the high non-thermal number density is robust and does not depend strongly on the particular choice of  $E_{\min}$ .

## A consistency check: comparison of microwave and EUV diagnostics of the coronal thermal plasma

A well-established way of investigating thermal coronal plasma is using extreme ultraviolet (EUV) emission, which is a combination of line emission

from ions, primarily iron, in various ionization states (and, thus, is temperature-sensitive) and a continuum owing to bremsstrahlung. Here we use EUV data taken by the Solar Dynamics Observatory Atmospheric Imaging Assembly (SDO/AIA) in six narrow passbands sensitive to EUV emission from the corona. For each pixel within the field of view that we used to analyse the microwave emission, we applied a regularized differential emission measure (DEM) inversion<sup>27</sup> technique, from which we derived the emission measure ( $\langle \langle \text{rm}\{\text{EM}\} \rangle \rangle = \langle \int \rangle \langle \langle \text{rm}\{\text{LOS}\} \rangle \rangle$ )  $\langle n \rangle \langle \langle \text{rm}\{\text{th}\} \rangle \rangle^2 \langle \langle \text{rm}\{\text{d}\} \rangle \rangle L$ , in which  $dL$  is the differential column depth along the LOS) as a moment of the DEM. The thermal number density is then estimated as  $\langle \langle n \rangle \rangle \langle \langle \text{rm}\{\text{th}\} \rangle \rangle = \sqrt{\langle \langle \text{rm}\{\text{EM}\} \rangle \rangle / L}$ , in which  $L$  is 5.8 Mm, as adopted for the microwave spectral model fitting. The EM distribution is shown in Extended Data Fig. 5a. Owing to rather strong EUV emission, the EM map contains saturated areas and diffraction artefacts. Therefore, for quantitative analysis, we selected a small rectangular area within ROI2 that avoids these artefacts to the extent possible.

Direct pixel-to-pixel comparison, even in the case of a perfect co-alignment, would be inconclusive in our case for the following reasons: (1) the pixel sizes of the AIA and EOVSA maps are different (0.6" and 2", respectively); (2) the time cadence of data used for the analysis are different (12 s and 4 s, respectively). Therefore, we compare statistical distributions, rather than individual values, of the thermal electron number density obtained from these two different datasets.

We consider a single 12-s time range of the AIA data in a small rectangle area, marked in dark blue in Extended Data Fig. 5, free from strong artefacts, which contains 100 AIA pixels, and three 4-s time ranges of the EOVSA data in ROI2 that contains 49 pixels, giving a total of 147 measurements over the same 12-s time range. The standard DEM inversion techniques assume the so-called coronal elemental abundances, for which the Fe abundance is four times larger than in the photosphere. It was reported<sup>28,29</sup>, however, that—in flaring volumes—the abundance can be closer to the photospheric one, owing to the fact that the thermal plasma is mainly due to chromospheric evaporation of material with photospheric abundance initiated by the precipitation of flare-accelerated particles into the chromospheric footpoint. Therefore, we used the AIA thermal plasma diagnostics assuming alternately both the coronal and the photospheric

abundance. Another possible source of uncertainty of the EUV diagnostics is an assumption of ionization equilibrium, which can be strongly violated during non-equilibrium flaring conditions. In addition, the EUV diagnostics suffer more from potential contributions along a long LOS (owing to the dependence of the EM on the column depth) compared with the microwave diagnostics, which are restricted to the region inside the non-thermal gyrosynchrotron source only.

With all these reservations in mind, Extended Data Fig. 5b shows a histogram of the thermal number density from the described rectangular ROI assuming the coronal abundance in filled dark blue and the photospheric abundance in empty dark blue. The filled light blue histogram shows the distribution of the thermal electron number density obtained for the three time frames for the entirety of ROI2. These distributions agree with each other within a factor of two (less for the photospheric abundance case), confirming that the thermal electron number densities derived from the microwave diagnostics in ROI2, in which they are statistically well constrained, are consistent with the EUV-derived numbers. We cannot perform a similar exercise in ROI1 because the microwave diagnostics of  $n_{\text{th}}$  does not offer well-constrained values.

## Data availability

All original EOVSA data are maintained on the EOVSA website at <http://www.ovsa.njit.edu/>. Original EOVSA data used for this study are available at

[http://www.ovsa.njit.edu/fits/IDB/20170910/IDB20170910155625/](http://www.ovsa.njit.edu/fits>IDB/20170910/IDB20170910155625/). Fully processed EOVSA spectral imaging data in IDL save format can be downloaded from

[http://ovsa.njit.edu/publications/fleishman\\_ea\\_science\\_2019/data/](http://ovsa.njit.edu/publications/fleishman_ea_science_2019/data/).

## Code availability

All the codes we used in this study are based on publicly available software packages: GSFIT is available in the community-contributed SolarSoftWare repository, under the packages category, at

[www.lmsal.com/solarsoft/ssw/packages/gsfit/](http://www.lmsal.com/solarsoft/ssw/packages/gsfit/); the open-source MCMC code is documented in ref. <sup>26</sup>.

## References

1. Yokoyama, T. & Shibata, K. Magnetic reconnection as the origin of X-ray jets and H $\alpha$  surges on the Sun. *Nature* **375**, 42–44 (1995).
2. Fleishman, G. D. et al. Decay of the coronal magnetic field can release sufficient energy to power a solar flare. *Science* **367**, 278–280 (2020).
3. Krucker, S. et al. Measurements of the coronal acceleration region of a solar flare. *Astrophys. J.* **714**, 1108–1119 (2010).
4. Krucker, S. & Battaglia, M. Particle densities within the acceleration region of a solar flare. *Astrophys. J.* **780**, 107 (2014).
5. Kontar, E. P. et al. Deducing electron properties from hard X-ray observations. *Space Sci. Rev.* **159**, 301–355 (2011).
6. Holman, G. D. et al. Implications of X-ray observations for electron acceleration and propagation in solar flares. *Space Sci. Rev.* **159**, 107–166 (2011).
7. Warren, H. P. et al. Spectroscopic observations of current sheet formation and evolution. *Astrophys. J.* **854**, 122 (2018).
8. Gary, D. E. et al. Microwave and hard X-ray observations of the 2017 September 10 solar limb flare. *Astrophys. J.* **863**, 83 (2018).
9. Longcope, D., Unverferth, J., Klein, C., McCarthy, M. & Priest, E. Evidence for downflows in the narrow plasma sheet of 2017 September 10 and their significance for flare reconnection. *Astrophys. J.* **868**, 148 (2018).
10. Hayes, L. A. et al. Persistent quasi-periodic pulsations during a large X-class solar flare. *Astrophys. J.* **875**, 33 (2019).

11. Chen, B. et al. Measurement of magnetic field and relativistic electrons along a solar flare current sheet. *Nat. Astron.* **4**, 1140–1147 (2020).
12. Yu, S. et al. Magnetic reconnection during the post-impulsive phase of a long-duration solar flare: bidirectional outflows as a cause of microwave and X-ray bursts. *Astrophys. J.* **900**, 17 (2020).
13. French, R. J., Matthews, S. A., van Driel-Gesztelyi, L., Long, D. M. & Judge, P. G. Dynamics of late-stage reconnection in the 2017 September 10 solar flare. *Astrophys. J.* **900**, 192 (2020).
14. Reeves, K. K. et al. Hot plasma flows and oscillations in the loop-top region during the 2017 September 10 X8.2 solar flare. *Astrophys. J.* **905**, 165 (2020).
15. Dreicer, H. Electron and ion runaway in a fully ionized gas. I. *Phys. Rev.* **115**, 238–249 (1959).
16. Priest, E. R. & Forbes, T. G. The magnetic nature of solar flares. *Astron. Astrophys. Rev.* **10**, 313–377 (2002).
17. Drake, J. F., Arnold, H., Swisdak, M. & Dahlin, J. T. A computational model for exploring particle acceleration during reconnection in macroscale systems. *Phys. Plasmas* **26**, 012901 (2019).
18. Arnold, H. et al. Electron acceleration during macroscale magnetic reconnection. *Phys. Rev. Lett.* **126**, 135101 (2021).
19. Drake, J. F., Swisdak, M. & Fermo, R. The power-law spectra of energetic particles during multi-island magnetic reconnection. *Astrophys. J. Lett.* **763**, L5 (2013).
20. Dahlin, J. T., Drake, J. F. & Swisdak, M. The mechanisms of electron heating and acceleration during magnetic reconnection. *Phys. Plasmas* **21**, 092304 (2014).
21. Dahlin, J. T., Drake, J. F. & Swisdak, M. Electron acceleration in three-dimensional magnetic reconnection with a guide field. *Phys. Plasmas* **22**, 100704 (2015).

22. Dahlin, J. T. Prospectus on electron acceleration via magnetic reconnection. *Phys. Plasmas* **27**, 100601 (2020).
23. Kuridze, D. et al. Mapping the magnetic field of flare coronal loops. *Astrophys. J.* **874**, 126 (2019).
24. Fleishman, G. D. & Toptygin, I. N. *Cosmic Electrodynamics. Astrophysics and Space Science Library* Vol. 388 (Springer, 2013).
25. Nita, G. M., Hickish, J., MacMahon, D. & Gary, D. E. EOVA implementation of a spectral kurtosis correlator for transient detection and classification. *J. Astron. Instrum.* **5**, 1641009 (2016).
26. Foreman-Mackey, D. et al. emcee: the MCMC hammer. *Publ. Astron. Soc. Pac.* **125**, 306 (2013).
27. Hannah, I. G. & Kontar, E. P. Differential emission measures from the regularized inversion of Hinode and SDO data. *Astron. Astrophys.* **539**, A146 (2012).
28. Warren, H. P. Measurements of absolute abundances in solar flares. *Astrophys. J. Lett.* **786**, L2 (2014).
29. Doschek, G. A. et al. Photospheric and coronal abundances in an X8.3 class limb flare. *Astrophys. J.* **853**, 178 (2018).

## Acknowledgements

We thank the scientists and engineers who helped design and build the EOVA, especially G. Hurford, S. White, J. McTiernan, W. Grammer and K. Nelin. This work was supported in part by NSF grants AGS-2121632, AGS-1817277, AST-1910354, AGS-2130832, AGS-1654382 and AST-2108853 and NASA grants 80NSSC18K0667, 80NSSC20K0627, 80NSSC20K1318, 80NSSC21K0623, 80NSSC19K0068 and 80NSSC18K1128 to New Jersey Institute of Technology.

## Author information

## **Authors and Affiliations**

1. Center for Solar-Terrestrial Research, New Jersey Institute of Technology, Newark, NJ, USA

Gregory D. Fleishman, Gelu M. Nita, Bin Chen, Sijie Yu & Dale E. Gary

## **Contributions**

G.D.F. developed the model spectral fitting methodology, participated in model fitting and analysis of the results and wrote the draft manuscript. G.M.N. developed the GSFIT spectral fitting package and participated in model fitting and analysis of results. B.C. developed the microwave spectral imaging and self-calibration strategy and adopted the MCMC methodology to perform the testing of the model fitting validity. S.Y. implemented the code of the microwave imaging pipeline under the guidance of D.E.G. and B.C. and performed the consistency check between the microwave and EUV diagnostics. D.E.G. led the construction and commissioning of the EOVSA and developed the observational strategy and calibration for microwave spectroscopy. All authors discussed the interpretation of the data, contributed scientific results and helped prepare the paper.

## **Corresponding author**

Correspondence to [Gregory D. Fleishman](#).

## **Ethics declarations**

## **Competing interests**

The authors declare no competing interests.

## **Peer review**

## **Peer review information**

*Nature* thanks the anonymous reviewers for their contribution to the peer review of this work. [Peer reviewer reports](#) are available.

## Additional information

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

## Extended data figures and tables

### [Extended Data Fig. 1 Example spectra from various locations and associated model fits.](#)

The central panel (**e**) shows a reference microwave image at 9.92 GHz at 15:58 UT. ROI1 and ROI2 are shown in green and cyan, respectively. Plus signs, 1 to 8, indicate pixels, for which the observed spectra and model fits are shown in the remaining eight panels (**a–d, f–i**), each marked with P1 to P8. The asterisks with error bars show the data and the uncertainties at the 1-sigma level; the curves show the corresponding model spectral fits. The fluxes are given in solar flux units (sfu):  $1 \text{ sfu} = 10^4 \text{ Jy} = 10^{-22} \text{ W m}^{-2} \text{ Hz}^{-1} = 10^{-19} \text{ erg s}^{-1} \text{ cm}^{-2} \text{ Hz}^{-1}$ .

### [Extended Data Fig. 2 MCMC probability distributions of the fit parameters for an example pixel in ROI2.](#)

The results are for pixel P2 located at  $x = 977.0''$  and  $y = -137.8''$ , as marked in Extended Data Fig. 1. Solid black horizontal/vertical lines in each panel indicate the best-fit values from the GSFIT minimization. Dotted blue horizontal/vertical lines mark the median values of the MCMC probability distributions. Dashed lines in the histograms along the diagonal indicate  $\pm 1$ -sigma standard deviation of a given parameter. Off-diagonal panels show correlations between all possible pairs of parameters shown as 2D histograms of the probability distributions. The contours represent 39.3%, 60% and 80% of the maximum. The outer contour is selected to represent approximately the 1-sigma region of a 2D Gaussian distribution ( $1 - e^{-0.5}$ ).

## Extended Data Fig. 3 MCMC probability distributions of the fit parameters for an example pixel in ROI1.

The figure layout is identical to Extended Data Fig. 2 but showing the parameters for pixel P1 located at  $x = 987.0''$  and  $y = -143.8''$ , as marked in Extended Data Fig. 1.

## Extended Data Fig. 4 MCMC constrained maps of the thermal and suprathermal electron parameters.

**a**, Thermal electron number density  $n_{\text{th}}$ . For each pixel, the median value of the MCMC probability distribution is shown. **b**, Similar to **a** but for the non-thermal electron number density  $n_{\text{nth}}$ . **c**, Map of the ratio of the upper limit of the thermal number density  $(n_{\text{th}})^{\text{upper}}$  (defined as one  $\sigma$  above the median  $n_{\text{th}} + \sigma$ ) to the non-thermal number density  $n_{\text{nth}}$ . **d**, Similar to **a** but for the low-energy cut-off  $E_{\text{min}}$ .

## Extended Data Fig. 5 Comparison of thermal plasma density diagnostics.

**a**, EM map obtained at 15:58 UT over the temperature range 0.5–30 MK. **b**, Histogram of the distribution of the thermal electron number density  $n_{\text{th}}$  in ROI2 and the ROI obtained from microwave and EUV diagnostics, respectively.

## Supplementary information

### Peer Review File

### Supplementary Video 1

Evolving maps of the coronal parameters of the solar flare. This Supplementary Video demonstrates evolution of the thermal number density, non-thermal number density and spectral index  $\delta$  of the suprathermal

electron distribution over energy in the 10 September 2017 flare over four minutes. Each frame is separated by 4 s. The solid white contour outlines ROI1.

## **Supplementary Video 2**

Acceleration efficiency versus electron spectral index  $\delta$ . This Supplementary Video demonstrates how the proxy of the acceleration efficiency,  $r_B$  (panel **a**), and the spectral index  $\delta$  (panel **b**) evolve, and how they establish a correlation as time progresses (panel **c**). The video layout is the same as that of Fig. 3.

## **Supplementary Video 3**

Sensitivity of total intensity of the microwave emission to variation of source parameters. This video demonstrates how the total intensity (Stokes I) of the microwave emission varies when the parameters of the emission source change one by one. Note, in particular, the effect of  $n_{\text{th}}$  and  $E_{\text{min}}$ .

## **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

Fleishman, G.D., Nita, G.M., Chen, B. *et al.* Solar flare accelerates nearly all electrons in a large coronal volume. *Nature* **606**, 674–677 (2022).  
<https://doi.org/10.1038/s41586-022-04728-8>

- Received: 28 July 2021
- Accepted: 04 April 2022
- Published: 08 June 2022
- Issue Date: 23 June 2022
- DOI: <https://doi.org/10.1038/s41586-022-04728-8>

### Share this article

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

[Get shareable link](#)

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

[Copy to clipboard](#)

Provided by the Springer Nature SharedIt content-sharing initiative

---

This article was downloaded by **calibre** from <https://www.nature.com/articles/s41586-022-04728-8>

- Article
- Open Access
- [Published: 22 June 2022](#)

# Observation of a correlated free four-neutron system

- [M. Duer](#),
- [T. Aumann](#),
- [R. Gernhäuser](#),
- [V. Panin](#),
- [S. Paschalis](#),
- [D. M. Rossi](#),
- [N. L. Achouri](#),
- [D. Ahn](#),
- [H. Baba](#),
- [C. A. Bertulani](#),
- [M. Böhmer](#),
- [K. Boretzky](#),
- [C. Caesar](#),
- [N. Chiga](#),
- [A. Corsi](#),
- [D. Cortina-Gil](#),
- [C. A. Douma](#),
- [F. Dufter](#),
- [Z. Elekes](#),
- [J. Feng](#),
- [B. Fernández-Domínguez](#),
- [U. Forsberg](#),
- [N. Fukuda](#),
- [I. Gasparic](#),
- [Z. Ge](#),
- [J. M. Gheller](#),

- [J. Gibelin](#),
- [A. Gillibert](#),
- [K. I. Hahn](#),
- [Z. Halász](#),
- [M. N. Harakeh](#),
- [A. Hirayama](#),
- [M. Holl](#),
- [N. Inabe](#),
- [T. Isobe](#),
- [J. Kahlbow](#),
- [N. Kalantar-Nayestanaki](#),
- [D. Kim](#),
- [S. Kim](#),
- [T. Kobayashi](#),
- [Y. Kondo](#),
- [D. Körper](#),
- [P. Koseoglou](#),
- [Y. Kubota](#),
- [I. Kuti](#),
- [P. J. Li](#),
- [C. Lehr](#),
- [S. Lindberg](#),
- [Y. Liu](#),
- [F. M. Marqués](#),
- [S. Masuoka](#),
- [M. Matsumoto](#),
- [J. Mayer](#),
- [K. Miki](#),
- [B. Monteagudo](#),
- [T. Nakamura](#),
- [T. Nilsson](#),
- [A. Obertelli](#),
- [N. A. Orr](#),
- [H. Otsu](#),
- [S. Y. Park](#),
- [M. Parlog](#),
- [P. M. Potlog](#),

- [S. Reichert](#),
- [A. Revel](#),
- [A. T. Saito](#),
- [M. Sasano](#),
- [H. Scheit](#),
- [F. Schindler](#),
- [S. Shimoura](#),
- [H. Simon](#),
- [L. Stuhl](#),
- [H. Suzuki](#),
- [D. Symochko](#),
- [H. Takeda](#),
- [J. Tanaka](#),
- [Y. Togano](#),
- [T. Tomai](#),
- [H. T. Törnqvist](#),
- [J. Tscheuschner](#),
- [T. Uesaka](#),
- [V. Wagner](#),
- [H. Yamada](#),
- [B. Yang](#),
- [L. Yang](#),
- [Z. H. Yang](#),
- [M. Yasuda](#),
- [K. Yoneda](#),
- [L. Zanetti](#),
- [J. Zenihiro](#) &
- [M. V. Zhukov](#)

[Nature](#) **volume 606**, pages 678–682 (2022)

- 854 Accesses
- 188 Altmetric
- [Metrics details](#)

# Abstract

A long-standing question in nuclear physics is whether chargeless nuclear systems can exist. To our knowledge, only neutron stars represent near-pure neutron systems, where neutrons are squeezed together by the gravitational force to very high densities. The experimental search for isolated multi-neutron systems has been an ongoing quest for several decades<sup>1</sup>, with a particular focus on the four-neutron system called the tetraneutron, resulting in only a few indications of its existence so far<sup>2,3,4</sup>, leaving the tetraneutron an elusive nuclear system for six decades. Here we report on the observation of a resonance-like structure near threshold in the four-neutron system that is consistent with a quasi-bound tetraneutron state existing for a very short time. The measured energy and width of this state provide a key benchmark for our understanding of the nuclear force. The use of an experimental approach based on a knockout reaction at large momentum transfer with a radioactive high-energy  ${}^8\text{He}$  beam was key.

# Main

A neutron can be bound either in an atomic nucleus or in a neutron star. The free neutron has a lifetime of just under 15 min and decays into a proton, electron and antineutrino. The system made of two neutrons, the dineutron, is known to be unbound by only about 100 keV. Whether multi-neutron systems can exist as weakly bound states or very short-lived unbound resonant states has been a long-standing question<sup>1</sup>. The next simplest system of three neutrons is less likely to exist owing to the odd number of nucleons and therefore weaker binding; yet, a recent calculation has suggested its existence<sup>5</sup>. Following these considerations, the four-neutron system, the tetraneutron, is an appropriate candidate to address this question. An overview of previous experiments and theoretical approaches is given in ref. <sup>1</sup>.

Numerous attempts have been made to find a hint for the existence of the tetraneutron as a bound or resonant state. Among these attempts, experiments were performed searching for possible bound tetraneutrons produced in uranium fission reactions (see, for example, ref. <sup>6</sup>). Other

attempts, sensitive to both bound and resonant states, used pion-induced double-charge-exchange (DCX) reactions, mainly the  $\langle\{\}^4\{\rm H\}\rangle\langle\{\rm e\}\rangle(\{\{\rm pi\}\}^{-},\{\{\rm pi\}\}^{+})$  reaction (see, for example, ref. <sup>7</sup>), as well as transfer reactions such as  $\langle\{\}^8\{\rm H\}\rangle\langle\{\rm e\}\rangle(\{\rm d\},\{\}^6\{\rm L\}\{\rm i\})$  (ref. <sup>8</sup>). None of the experiments yielded a positive signal.

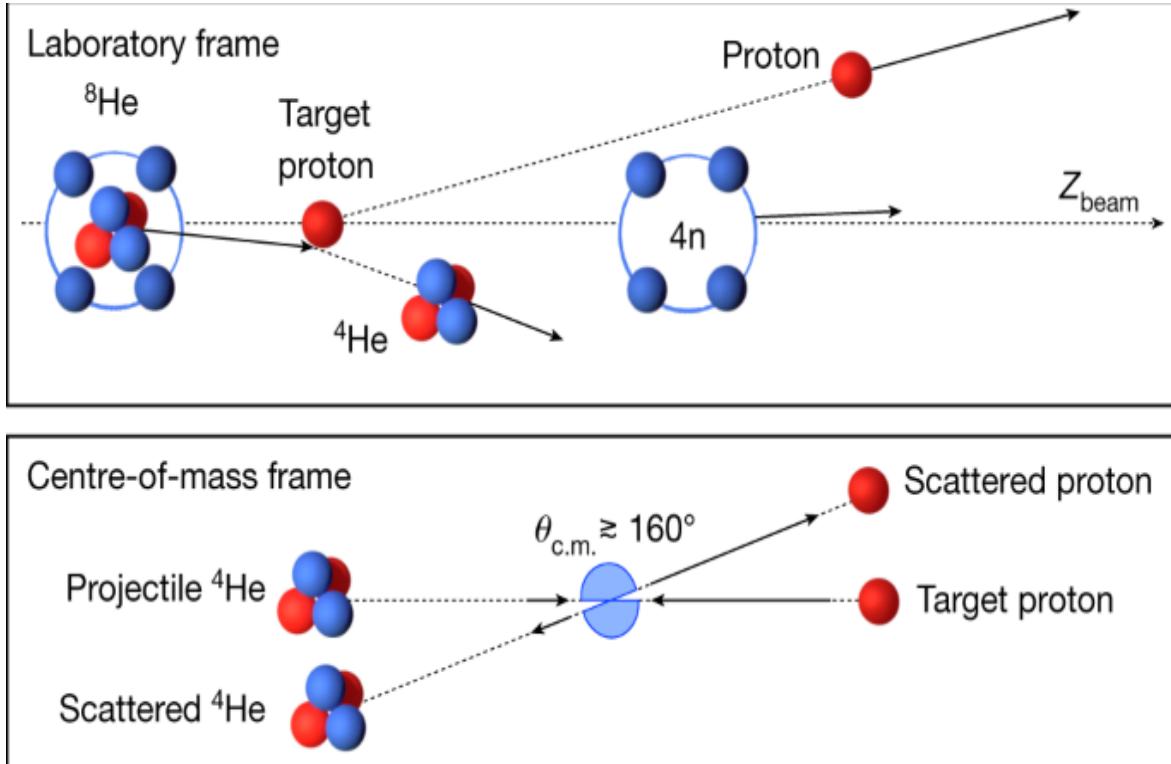
Most of the past experiments were performed with stable nuclei. Towards the twenty-first century, with the development of radioactive-ion beam facilities, it became possible to use extremely neutron-rich nuclei in which one can expect an enhanced formation of a tetraneutron system. The first indication for a possible bound tetraneutron was reported in 2002<sup>2</sup> from a break-up reaction of <sup>14</sup>Be into <sup>10</sup>Be + <sup>4</sup>n. The result stimulated several theoretical studies, all agreeing on the same conclusion: a bound tetraneutron state cannot be obtained theoretically without significantly changing our understanding of the nuclear forces<sup>9,10,11</sup>. However, the possibility of the four-neutron system existing as a resonant quasi-bound state with a very short lifetime on the order of a few  $10^{-22}$  s, before decaying, has remained an open and challenging question. It was later found that the result reported in ref. <sup>2</sup> is also consistent with such a resonant state with the limit on its energy  $\langle\{E\}_{\{\rm r\}}\rangle \leq 2\,\{\rm MeV\}$  (ref. <sup>3</sup>).

A decade later, in 2016, an indication of a tetraneutron resonance was reported<sup>4</sup>. A DCX reaction was used, but in contrast to previous attempts, this time the reaction was induced by a high-energy <sup>8</sup>He radioactive beam. <sup>8</sup>He is the most neutron-rich bound isotope, and the  ${}^8\text{He}({}^4\text{He}, {}^8\text{Be})$  reaction channel was investigated. The advantage of using a radioactive beam is the freedom of selecting the reaction partner in a so-called recoil-less production (without momentum transfer) of the four-neutron system. The energy of the state was found to be  $E_r = 0.8 \pm 1.4$  MeV, and an upper limit on its width was estimated as  $\Gamma \leq 2.6$  MeV. However, owing to the large experimental uncertainty, the possibility of a bound state could not be excluded by this experiment.

In this work, we used the quasi-elastic knockout of an  $\alpha$ -particle (<sup>4</sup>He nucleus) from a high-energy <sup>8</sup>He projectile induced by a proton target to

populate a possible tetraneutron state. The inverse-kinematics knockout reaction  $\left(\{ \}^8 \{ \text{rm}\{\text{He}\} \} (\{ \text{rm}\{\text{p}\} \}, \{ \{ \text{rm}\{\text{p}\} \}^4 \{ \text{rm}\{\text{He}\} \})\right)$  at large momentum transfer is well suited because the  ${}^8\text{He}$  nucleus has the pronounced cluster structure of an  $\alpha$ -core ( ${}^4\text{He}$ ) and four valence neutrons with small  $4n$  centre-of-mass motion, such that after the sudden removal of the  $\alpha$ -particle, a rather localized four-neutron system with small relative energy between the neutrons is produced, which may have a large overlap with a tetraneutron state<sup>12,13</sup>. The chosen kinematics at large momentum transfer between the proton and the  $\alpha$ -particle ensures that the four-neutron system will recoil only with the intrinsic momentum of the  ${}^4\text{He}$  core in the  ${}^8\text{He}$  rest frame, without any further momentum transfer, thus allowing the recoil-less production. Furthermore, final-state interactions between the four neutrons and the charged particles are also minimized owing to the large momentum transfer, separating charged reaction partners from the neutron spectators in momentum space (Fig. 1).

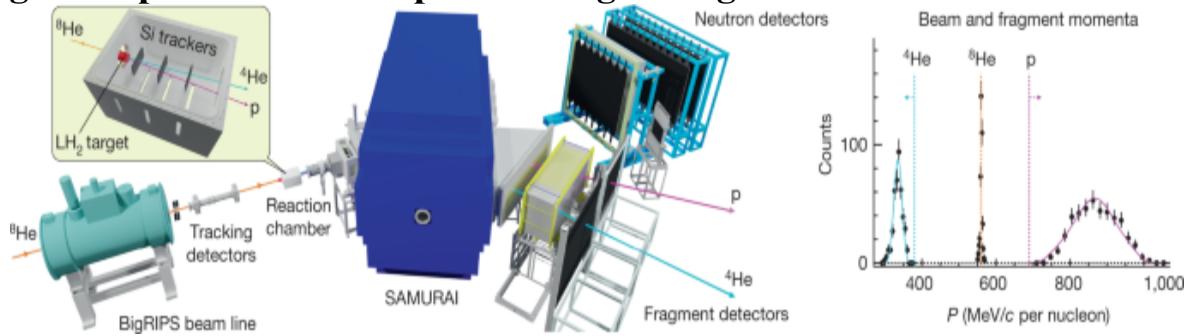
**Fig. 1: Schematic illustration of the quasi-elastic reaction investigated in this work.**



Top: quasi-elastic scattering of the  ${}^4\text{He}$  core from a  ${}^8\text{He}$  projectile off a proton target in the laboratory frame. The length of the arrows represents the momentum per nucleon (the velocity) of the incoming and outgoing particles.  $Z_{\text{beam}}$  is the beam axis. Bottom: the equivalent p- ${}^4\text{He}$  elastic scattering in their centre-of-mass frame, where we consider reactions at backward angles close to  $180^\circ$ ,  $\theta_{\text{c.m.}} \gtrsim 160^\circ$ . In this frame, the momentum of the proton balances that of the  ${}^4\text{He}$ ,  $\langle \{\{\text{bf}\{P\}\}\} \rangle - \langle \{\{\text{rm}\{p\}\}\} \rangle = \langle \{\{\text{bf}\{P\}\}\} \rangle / \langle \{\{\text{rm}\{He\}\}\} \rangle$ , that is, the proton is four times faster than the  ${}^4\text{He}$ .

The experiment took place at the Radioactive Ion Beam Factory operated by the RIKEN Nishina Center and the Center for Nuclear Study, University of Tokyo, using the Superconducting Analyzer for Multi-particles from Radio Isotope Beams (SAMURAI)<sup>14</sup>. A primary beam of  ${}^{18}\text{O}$  was directed onto a beryllium production target producing a cocktail of radioactive nuclei from fragmentation. The secondary  ${}^8\text{He}$  beam was separated using the BigRIPS fragment separator and transported with an energy of 156 MeV per nucleon to a 5-cm-thick liquid-hydrogen target<sup>15</sup> located at the SAMURAI spectrometer (Fig. 2).

**Fig. 2: Experimental set-up and charged fragments momenta.**



Left: schematic view of the experimental set-up. The  ${}^8\text{He}$  secondary beam at 156 MeV per nucleon is transported from the BigRIPS (Big RIKEN projectile-fragment separator) into the SAMURAI set-up, where it hits a liquid-hydrogen ( $\text{LH}_2$ ) target. In a quasi-elastic  $\langle \{\{\text{rm}\{p\}\}\}, \langle \{\{\text{rm}\{p\}\}\}^4 \rangle \rangle$  reaction, the  ${}^4\text{He}$  core is knocked out from the  ${}^8\text{He}$  projectile. Scintillator detectors and drift chambers are used for beam identification and tracking. The trajectories of the outgoing fragments are

tracked by three silicon (Si) planes and bent afterwards through the SAMURAI spectrometer towards the focal-plane detectors. Two neutron-detector arrays were placed at a forward angle behind the SAMURAI. An additional scintillator wall was placed at smaller bending angle to detect the unreacted  $^8\text{He}$  beam. Right: measured momenta of the knocked-out  $^4\text{He}$  and the scattered proton after the quasi-elastic scattering (symbols). The momentum distribution of the incoming  $^8\text{He}$  beam is shown for comparison. The solid curves are the results from the simulation. The cyan (magenta) dotted line represents the upper (lower) limit of the  $^4\text{He}$  (proton) momentum expected from the simulation assuming a quasi-elastic scattering, and the orange line indicates the central beam momentum.

### Source data

The incoming beam was measured upstream of the target on an event-by-event basis using scintillators for charge identification as well as momentum measurement, and two drift chambers for tracking (Extended Data Fig. 1).

The outgoing charged fragments ( $\alpha$ -particle and proton) emerging from the quasi-elastic scattering were detected using a combination of detectors downstream of the target. Three planes of silicon-strip detectors, where each plane consists of two orthogonal layers enabling position measurements in both horizontal and vertical directions, served for tracking, energy-deposition measurement and reconstruction of the reaction vertex inside the target (Extended Data Figs. 2 and 3).

Behind the silicon planes, both charged fragments were bent through the magnetic field of the SAMURAI spectrometer, which was operated at a nominal magnetic field of 1.25 T in the centre of the magnet. The experiment was designed to detect an  $\alpha$ -particle and a proton that emerge from quasi-elastic scattering close to  $180^\circ$  in the centre-of-mass frame (Fig. 1). Under these kinematical conditions, their resulting outgoing momenta are very different from each other in the laboratory frame, as shown in Figs. 1 and 2. The knocked-out  $\alpha$ -particle is slowed down from its initial momentum, that is, with the incoming beam velocity, to a momentum of about  $330 \text{ MeV}/c$  per nucleon after the reaction (where  $c$  is the speed of light). In contrast, the proton, which was at rest in the initial state, becomes the fastest particle in the reaction, gaining a typical momentum of about  $860$

MeV/c. At the focal plane, a drift chamber is used for tracking of the fragments after the magnet, and two scintillator walls located side by side, which cover a wide momentum range, are used for energy-deposition and time-of-flight measurements. The  $\alpha$ -particle and proton are identified from a combination of their measured energy deposition, each in a different scintillator wall, and their position in the drift chamber (Extended Data Fig. 4). Their momenta are determined precisely from their reconstructed trajectories through the SAMURAI spectrometer.

As no additional momentum is transferred to the neutrons in the reaction, they continue moving with nearly beam velocity and can be detected, in principle, by the neutron detectors placed at a forward angle behind the SAMURAI spectrometer. The detection efficiency for neutrons is much lower than that for charged particles and decreases quickly as a function of the number of detected neutrons. The small p– ${}^4\text{He}$  elastic cross-section at backwards centre-of-mass angles of less than 1 microbarn (ref. 16) resulted in the relatively low statistics of 422 events obtained for the  $\backslash(\{\}^{\wedge}\{8\}\{\text{rm}\{\text{H}\}\}\{\text{rm}\{\text{e}\}\})(\{\text{rm}\{\text{p}\}\},\,\{\{\text{rm}\{\text{p}\}\}\}^{\wedge}\{4\}\{\text{rm}\{\text{H}\}\}\{\text{rm}\{\text{e}\}\})\backslash$  reaction. These factors made it impossible to detect more than two neutrons in coincidence with the charged particles. Therefore, the neutron detection is not a part of the current study, aside from a consistency check (provided in [Supplementary Information](#)) of the near recoil-less production of the free neutrons.

The combined selection of event-by-event identification of incoming  ${}^8\text{He}$ -beam particles in coincidence with the knocked-out  $\alpha$ -particle and the scattered proton defines the  $\backslash(\{\}^{\wedge}\{8\}\{\text{rm}\{\text{H}\}\}\{\text{rm}\{\text{e}\}\})(\{\text{rm}\{\text{p}\}\},\,\{\{\text{rm}\{\text{p}\}\}\}^{\wedge}\{4\}\{\text{rm}\{\text{H}\}\}\{\text{rm}\{\text{e}\}\})\backslash$  channel. From a precise measurement of the momenta of the charged particles, the energy spectrum of the 4n system is reconstructed assuming energy and momentum conservation through the missing mass:

$$\$\$ \{E\}_{\text{miss}} = \sqrt{\{E\}_{\text{miss}}^2 - \{P\}_{\text{miss}}^2} - 4m_n, \$\$$$

(1)

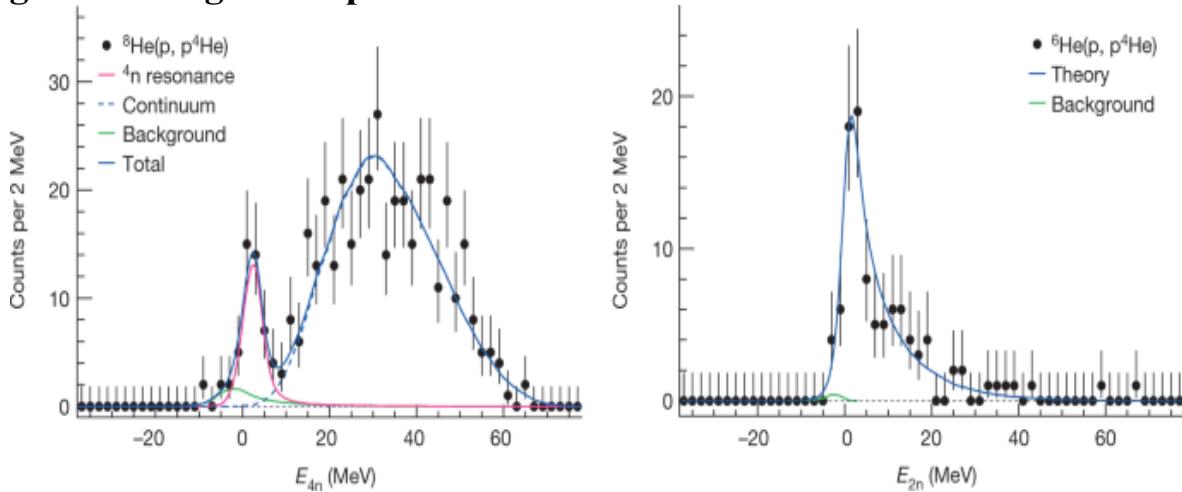
where  $E_{\text{miss}}$  ( $\mathbf{P}_{\text{miss}}$ ) is the energy (momentum) component of the missing-momentum four-vector, and  $m_n$  is the neutron mass. Using this notation, a bound  $^4\text{n}$  system will appear at  $E_{4n} < 0$  whereas a resonant state will appear at  $E_{4n} > 0$ . The missing momentum in equation (1) is defined by \

$$(\{\bar{P}\}_- \{\rm miss\}) = (\bar{P}_- \{} \{}^8 \{\rm He\} \} + \\ (\bar{P}_- \{\rm p\} (\{\rm tgt\})) - (\bar{P}_- \{} \{}^4 \{\rm He\} \} - \\ (\bar{P}_- \{\rm p\} \}), \text{ where the four-momenta } (\bar{P}_-) \text{ on the right-} \\ \text{hand side of the equation are those of the incoming beam, target proton,} \\ \text{knocked-out } \alpha\text{-particle and scattered proton, respectively.}$$

The  $(\{} \{}^6 \{\rm H\} \{\rm e\} (\{\rm p\}, \{\rm p\})^4 \{\rm H\} \{\rm e\})$  knockout reaction was measured with almost exactly the same experimental conditions as for  $^8\text{He}$ , except for some small differences in the energy of the incoming beam and the beam profile ([Supplementary Table 2](#)), and served as a benchmark for verifying the analysis and calibration procedures. In the case of  $^6\text{He}$ , the  $2\text{n}$  system is produced by the sudden removal of the  $^4\text{He}$  core. The two-neutron relative-energy spectrum is expected to be well described by theory taking into account both the well established ground-state wavefunction and the final-state scattering wave of the two neutrons, predicting a low-energy peak around 100 keV. Similarly to the  $^8\text{He}$  case, we define the missing mass  $((\bar{P}_- \{} \{}^8 \{\rm He\} \}) \rightarrow (\bar{P}_- \{} \{}^6 \{\rm He\} \})$  and  $(4m_- \{\rm n\}) \rightarrow (2m_- \{\rm n\})$ . The measured missing-mass spectrum for  $^6\text{He}$  is shown in the right panel of Fig. 3 together with the theoretical calculation<sup>17</sup> convoluted with the experimental acceptance and resolution (blue curve). The energy range shown represents the one covered by the experimental acceptance. The calculation is compared with the data by implementing it into an event generator for the quasi-elastic reaction, which uses the measured  $\text{p}-^4\text{He}$  differential elastic cross-section<sup>16</sup> as an input, as well as the measured internal momentum distribution of the  $\alpha$ -particle in  $^6\text{He}$  (ref. [18](#)). The generated events are transported through the experimental set-up in Geant4 simulations to account for the experimental acceptance and detector resolutions. The excellent agreement of the simulated theoretical distribution with the measured spectrum confirms the analysis and the calibration for determining the missing mass. The missing-mass resolution obtained in the measurement is approximately 1 MeV

sigma, and is almost constant over the measured energy range. The systematic uncertainty for the determination of the absolute energy was estimated from this measurement to be 0.4 MeV and that of the energy width to be 0.27 MeV ([Methods](#)). Also shown in the right panel of Fig. 3 (green curve) is a possible small background contribution coming from two-step process where  ${}^4\text{He}$  is produced in a first step (see [Methods](#) and following discussion for  ${}^8\text{He}$ ). This background was estimated from the measured cross-section to contribute 1% of the total number of measured events.

**Fig. 3: Missing-mass spectra.**



Left: missing-mass spectrum of the four-neutron system extracted from the  $\langle \langle \rangle \rangle^8 \{\text{rm}\{\text{He}\}\}(\{\text{rm}\{\text{p}\}\}, \langle \langle \rangle \rangle^4 \{\text{rm}\{\text{He}\}\})$  reaction. The different curves represent a Breit–Wigner resonance (pink), a non-resonant continuum (dashed blue), the background from two-step processes (green) and the total sum (solid blue). Right: missing-mass spectrum of the two-neutron system extracted from the  $\langle \langle \rangle \rangle^6 \{\text{rm}\{\text{He}\}\}(\{\text{rm}\{\text{p}\}\}, \langle \langle \rangle \rangle^4 \{\text{rm}\{\text{He}\}\})$  reaction. The blue curve represents the theoretical calculation<sup>17</sup> convoluted with the experimental acceptance and resolution, and the green curve represents the background from the two-step reaction.

### [Source data](#)

The measured missing-mass spectrum of the four-neutron system from the  $\langle \langle \rangle \rangle^8 \{\text{rm}\{\text{He}\}\}(\{\text{rm}\{\text{p}\}\}, \langle \langle \rangle \rangle^4 \{\text{rm}\{\text{H}\}\} \{\text{rm}\{\text{e}\}\})$  reaction is shown in the left panel of Fig. 3. Two components are observed: a

well pronounced peak in the low-energy region with an energy around 2 MeV and a broad distribution at higher energies attributed to a non-resonant continuum response<sup>13</sup>, a direct four-body decay.

The shape of the non-resonant continuum spectrum of the four neutrons has been studied theoretically for the case of the four-neutron structure formed after the sudden removal of the  $\alpha$ -core from  $^8\text{He}$  (ref. 13). The creation of the system is investigated by introducing into the Schrödinger equation a source term that accounts for the reaction mechanism producing the four-body system, and that depends explicitly on the internal structure of the parent nucleus  $^8\text{He}$ . The  $^8\text{He}$  ground-state wavefunction (without final-state interaction) was treated using the five-body ( $\{\}^4\{\}\{\rm He\}+4\{\rm n\}\{\}$ ) cluster orbital shell model approximation (COSMA)<sup>12</sup>. The exact shape of the non-resonant continuum is sensitive to the hyperradius of the source,  $\rho_{\text{sour}}$  an internal radius of the 4n system, described in the hyperspherical harmonics basis. A hyperradius of 5.6 fm is considered by the theory as the most realistic, as it reproduces the correct experimental radius of  $^8\text{He}$  in the COSMA model. This results in a broad distribution centred around 30 MeV, in good agreement with the observed experimental spectrum.

We model the spectrum as follows:

$$\begin{aligned} \$\$f(\{E\}_\{4\{\rm n\}\}) = & a\{f\}_\{\{\rm res\}\} \\ (\{E\}_\{4\{\rm n\}\}) + & b\{f\}_\{\{\rm con\}\} \\ (\{E\}_\{4\{\rm n\}\}) + & c\{f\}_\{\{\rm bkg\}\}(\{E\}_\{4\{\rm n\}\}), \$\$ \end{aligned} \quad (2)$$

where  $a$ ,  $b$ , and  $c$  are constants,  $f_{\text{res}}$  is a Breit–Wigner function representing the possible resonance structure, and  $f_{\text{con}}$  is the non-resonant continuum part presented above with the hyperradius as a parameter. The last term in equation (2),  $f_{\text{bkg}}$ , represents possible background events coming from competing processes. Several processes were investigated and quantified ([Methods](#)), where the only non-negligible contribution found is from a two-step process involving  $^6\text{He}$  ( $^4\text{He}$ ) production: proton-induced break-up of  $^8\text{He}$  into  $^6\text{He}$  ( $^4\text{He}$ ) followed by a p– $^4\text{He}$  quasi-elastic scattering. The resulting energy distribution is broadened and shifted to lower energies

compared with the pure  $^6\text{He}$  case (right panel of Fig. 3) owing to the two-step process, which has been taken into account in the simulation of  $f_{\text{bkg}}$ . This background was estimated from measured cross-sections to contribute 2.6% to the total number of measured events ([Methods](#)), which has been used to determine the normalization constant  $c$ .

The experimental spectrum was then fitted with the energy-dependent function given in equation (2), where the fit function was convoluted with the experimental response, taking into account acceptance and detector resolutions. The experimental acceptance is not constant over the measured energy range. It is maximal in the region  $\langle 20, \text{MeV} \rangle < \langle E \rangle_{\{4\}} < 40, \text{MeV} \rangle$  (Extended Data Fig. 5).

The result of the  $\chi^2$  minimization is presented by the solid blue curve in the left panel of Fig. 3, together with the individual contributions. The statistical significance of the peak structure is well beyond the  $5\sigma$  level ([Methods](#)).

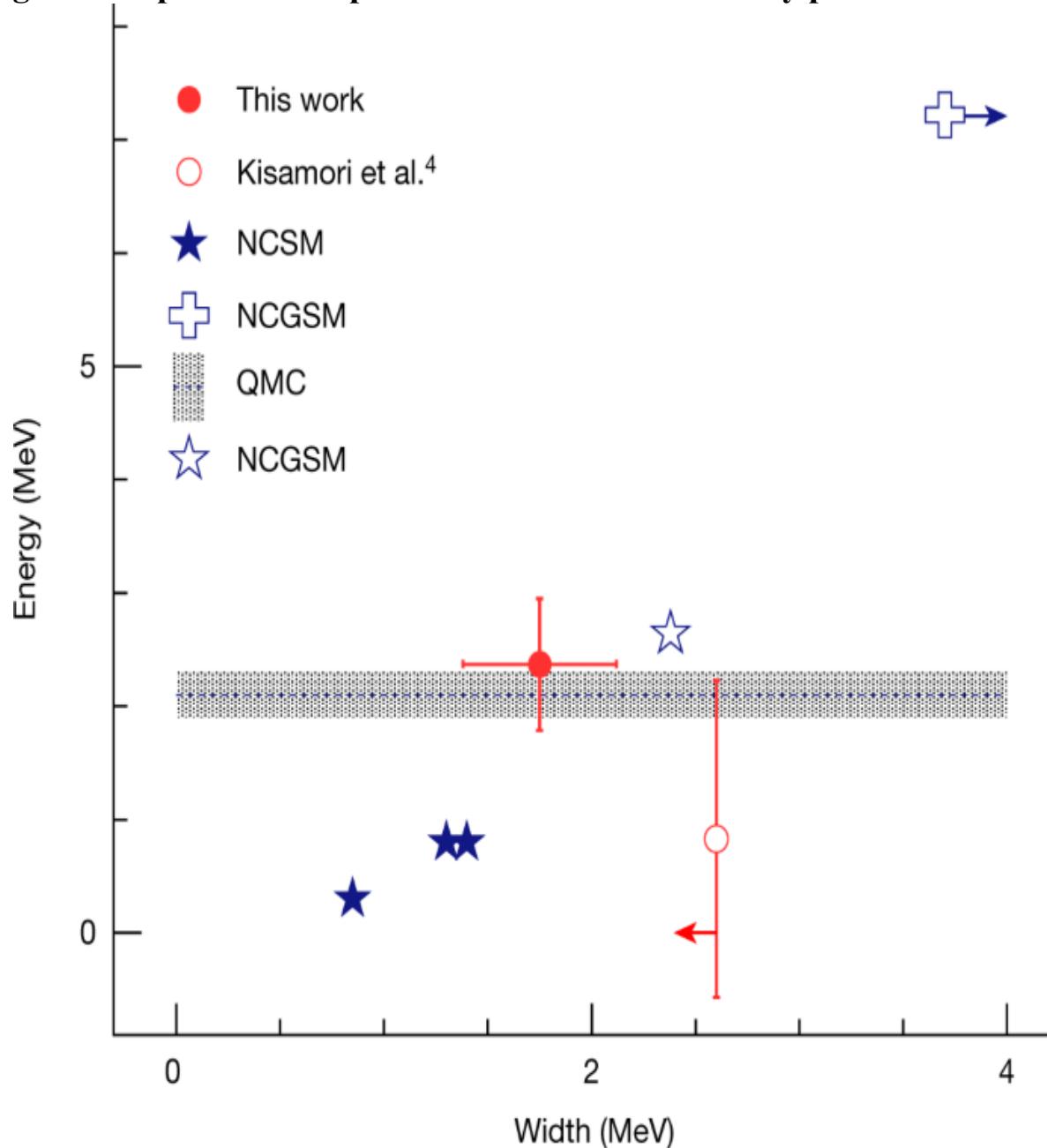
The probability of populating a four-neutron system in a resonant state after the sudden removal of the  $\alpha$ -core in  $^8\text{He}$  is determined by the overlap of the  $4n$  wavefunction in the final state and the relative motion of the four neutrons in the  $^8\text{He}$  initial state. This overlap integral defines the ratio of cross-sections for the population of the resonance and the non-resonant continuum. Unconvoluting with the acceptance of the set-up, following the energy dependence of equation (2), we extract a probability of  $P_r = 18.7 \pm 2.3\%$ . For comparison, the relative motion of the four neutrons studied in the COSMA model<sup>[12,13](#)</sup> yields a probability of about 30%. This value is obtained by considering the hyperradius of 5.6 fm, whereas the resulting value from the fit to the experimental data is  $5.0 \pm 0.1$  fm, which would yield a smaller probability to populate the resonant state.

Assuming a resonant state, its energy and width as determined from the fit are

$$\begin{array}{c} E_r = 2.37 \pm 0.38(\text{stat}) \pm 0.44(\text{sys.}), \text{ MeV}, \\ \varGamma = 1.75 \pm 0.22(\text{stat}) \pm 0.30(\text{sys.}), \text{ fm} \end{array}$$

For comparison, Fig. 4 shows our result (full red symbol) together with the previous experimental result obtained from the DCX measurement<sup>4</sup> (open red symbol). The energy of the resonance is in agreement within the uncertainty, despite the fact that different reactions were used to probe the 4n system, and is also in agreement with the upper limit given in ref. <sup>3</sup>.

**Fig. 4: Comparison of experimental results with theory predictions.**



Energy versus width of a tetraneutron resonance. Experimental data are shown in red: this work (full symbol), and the result from the DCX measurement<sup>4</sup> (open symbol), where the red arrow indicates that the measured width is an upper limit. Theory predictions are shown in blue based on: NCSM<sup>19</sup> and ref. 20 cited in ref. <sup>20</sup> (full stars), NCGSM (open star)<sup>20</sup> (cross)<sup>21</sup>, where the blue arrow indicates that the width is predicted to be larger than 3.7 MeV, and QMC calculations<sup>5</sup> (band). Whether this observation of a low-energy peak is attributed to a four-neutron resonant state or to other correlations between the neutrons in the final state, needs to be clarified by ab initio theories.

### Source data

From the theory side, there is no consensus among the different theories and their predictions are partly contradictory, although, there is a general agreement that a bound  $^4n$  does not exist. In 2003, Pieper<sup>10</sup> studied this possibility using Green's function Monte Carlo calculations. His conclusion was that the existence of a bound  $^4n$  state has to be excluded, unless nuclear forces are drastically modified. However, his calculation suggested that a possible resonance might exist near 2 MeV, but in such a case it must be very broad.

Using a similar approach, the quantum Monte Carlo (QMC) framework based on two-body and three-body chiral interactions was used to calculate the energy of a  $^4n$  resonance<sup>5</sup>. The result supports the existence of a resonant state with an energy of 2.1(2) MeV, while no prediction has been made for its width (blue band). An extended no-core shell model (NCSM) approach using a harmonic-oscillator representation predicts different resonant states (full stars) including their corresponding widths<sup>19</sup> (see also ref. <sup>20</sup> cited in ref. <sup>20</sup>). Calculations have been performed also in the framework of the no-core Gamow shell model (NCGSM)<sup>21</sup>. These resulted in  $E_r \approx 7$  MeV and  $\Gamma \approx 3.7$  MeV (cross), where the conclusion was in fact that the energy of a  $^4n$  resonance might be compatible with the experimental value of ref. <sup>4</sup>, albeit with a significantly larger width. As pointed out in a later study<sup>20</sup>, these calculations were incomplete, as they were performed only in truncated model spaces or with unphysically overbinding interactions. The authors of

this work<sup>20</sup> concluded that both the energy and the width of such a resonance are comparable with the experimental data (open star). At the same time, other calculations claim that to generate such a resonance, nuclear forces have to be significantly changed<sup>22,23,24,25,26,27</sup>, which would not be consistent with the present understanding. We note that some theories<sup>26,27</sup> predict a non-resonant low-energy enhancement of the density of states in the four-neutron spectrum. Whether such a prediction is consistent with our observed resonance-like feature cannot be currently ascertained, as the energy spectrum of the four-neutron system is not given. The drastically different predictions resulting from different theoretical approaches highlight the importance of the current firm experimental observation.

In conclusion, we have presented the experimental observation of a resonance-like structure consistent with a tetraneutron state near threshold after 60 years of experimental attempts to clarify the existence of this state. The use of a high-energy knockout reaction in inverse kinematics allowed a precise measurement. The use of a radioactive  ${}^8\text{He}$  beam as the parent system and a direct, large momentum-transfer reaction opened up the opportunity to create the  $4n$  system in a well defined one-step process and in a recoil-less undisturbed way. The optimized detection system enabled a precise determination of the final state and a high-resolution measurement. A well developed peak structure has been observed at an energy of  $2.37 \pm 0.38(\text{stat.}) \pm 0.44(\text{sys.})$  MeV with a striking statistical level. This is in agreement with the result of ref. <sup>4</sup> and the upper limit given in ref. <sup>3</sup>. Both the energy and the extracted width of  $\Gamma = 1.75 \pm 0.22(\text{stat.}) \pm 0.30(\text{sys.})$  MeV are consistent with a tetraneutron state that is unbound with a corresponding lifetime of  $(3.8 \pm 0.8) \times 10^{-22}$  s. Next-generation experiments using different reaction mechanisms and possibly detecting the four neutrons in coincidence will reveal more insights into the properties of the four-neutron system, including correlations among the neutrons. More elaborated ab initio nuclear theories accounting fully for the effect of the continuum are necessary to understand the observed low-energy peak and its origin.

## Methods

## Energy $E_{2n}$ of the $2n$ system

The theoretical calculation used as an input for the  $E_{2n}$  distribution includes both the ground-state wavefunction and the final-state interaction between the two neutrons<sup>17</sup>. These were obtained from a three-body ( $\{\}^4 \{\rm He\} + 2 \{\rm n\}$ ) cluster model, which contains a phenomenological three-body force as well as the following local angular momentum  $l$ -dependent two-body interactions: a central  $s$ -wave interaction in the  $nn$  system, and central  $s$ -,  $p$ - and  $d$ -wave interactions in the  $n-\alpha$  system, as well as  $n-\alpha$  spin-orbit interactions on the  $p$ - and  $d$ -waves. The ground-state wavefunction is computed in a three-body model using the computer code FaCE<sup>28</sup>, an acronym for Faddeev with Core Excitation, whereas the final-state interaction is taken into account by using the  $nn$  transition  $t$ -matrix approach. The calculated distribution is shown in Extended Data Fig. 6 (purple). For comparison, we also show the calculated distribution without the  $nn$  final-state interaction (green), which reflects the  $nn$  motion in the source, the  ${}^6\text{He}$  ground state. The distributions are normalized to the same maximum, such that only their shape is compared. It can be seen that the effect of the  $nn$  final-state interaction is very large. The ground-state distribution peaks at a larger energy with a far extending tail towards higher energies.

## Systematic uncertainty of the missing-mass determination

As the theoretical calculation used as an input for the energy of the two-neutron system in the quasi-elastic  $\{\}^6 \{\rm He\} (\{\rm p\}, \{\rm p\}) \{\}^4 \{\rm He\})$  reaction is considered to be accurate, the comparison of it with the measured spectrum is used to estimate the systematic uncertainty of the missing-mass measurement. Different offsets covering the range of  $\pm 0.5$  MeV are added to the generated theoretical distribution and for each offset the goodness of the agreement with the data is quantified by calculating the  $\chi^2$  value. The generated distributions are convoluted with the energy response of the set-up, to account for the experimental acceptance and the detector resolutions. The minimum  $\chi^2$  obtained is located very close to zero, that is, very close to the original distribution (Extended Data Fig. 7). The range of deviations in the energy scale still showing reasonable

agreement with the data ( $\langle \chi^2_{\text{min}} \rangle^{(2)} \pm 1$ ) range) is interpreted as the systematic uncertainty for the absolute missing-mass determination. This results in an uncertainty with an average value of  $\pm 0.4$  MeV. To determine the systematic uncertainty for the missing-mass width, the original theoretical distribution is used (that is, offset of zero). To check the sensitivity to the resolution, different resolutions are applied in the range  $(\sigma = 1 \pm 0.5, \text{MeV})$ . The generated distributions are convoluted with the acceptance of the set-up and are smeared using a Gaussian function with a width  $\sigma$ . Similarly as described above, the  $\chi^2$  values are calculated and the ( $\chi^2_{\text{min}} \rangle^{(2)} \pm 1$ ) range is interpreted as the systematic uncertainty (Extended Data Fig. 7). This results in an uncertainty with an average value of  $\pm 0.27$  MeV. We note that for both absolute value and resolution, the minimum  $\chi^2$  values fall very close to the expected values (which could be statistically by accident). As the given systematic uncertainties are determined solely by the statistics of the  ${}^6\text{He}$  data, we consider the given uncertainties as conservative estimates.

An additional systematic uncertainty from the background function in equation (2), on both the energy and the width, is added quadratically. The fitting procedure is repeated three times, with the green curve presented in Fig. 3 and with the  $1\sigma$  lower and upper limits (Supplementary Information). The values for the energy and the width are taken as the average between the different fits, and the systematic uncertainty is taken as the difference between them. This results in an uncertainty of 0.18 MeV and 0.14 MeV for the energy and width, respectively.

## Energy $E_{4n}$ of the 4n system

The relative-energy spectrum of four neutrons produced by sudden removal of the  $\alpha$ -particle from  ${}^8\text{He}$  as a source was studied theoretically in ref. 13. The final-state energy spectrum depends on the intrinsic 4n relative motion in  ${}^8\text{He}$ , that is, the  ${}^8\text{He}$  ground-state wavefunction as the source (without final-state interaction), and the final scattering state of the four isolated neutrons, that is, 4n final-state interaction. The wavefunction used in ref. 13 for  ${}^8\text{He}$  is based on the COSMA model<sup>12</sup>, and the source term contains the Fourier transform of the overlap between  ${}^8\text{He}$  and the  $\alpha$ -particle

wavefunctions. The method of hyperspherical harmonics<sup>29</sup> was used to solve the equations of the model. This method is based on the link between the hyperspherical-function method and the oscillator NCSM and uses a Slater determinant representation of the hyperspherical harmonics functions. The resulting source function is described by the internal variables of the four-neutron system, hyperradius, hyperangle, hyperangular momentum and the  ${}^4\text{He}$ -4n relative motion. Calculations showed that only hyperangular momentum  $K = 2$  significantly populates the 4n non-resonant continuum, reflecting the motion of the four neutrons in  ${}^8\text{He}$ . The resulting energy spectrum in ref. <sup>13</sup> is presented only up to 20 MeV. It was also used in ref. <sup>4</sup> to estimate the non-resonant background and can be modelled analytically as

$$\$ \$ \{f\}_{\{\{\rm cont\}\}}(\{E\}_{\{4\{\rm n\}\}}) = \{E\}_{\{4\{\rm n\}\}}^{\alpha} \times \exp(-\{E\}_{\{4\{\rm n\}\}}/\{\epsilon\}_{\{\{\rm a\}\}}), \$ \$$$

(3)

where  $\alpha = 7/2 + K$  with  $K = 2$  as stated above.  $K = 0$  and  $\alpha = 7/2$  corresponds to the free four-body phase space.  $\sqrt{\{\epsilon\}_{\{\{\rm a\}\}}}$  approx  $\frac{3.3\{\hbar\}^2}{\{m_N\}\{\rho_{\rm sour}\}^2}$ , where  $m_N$  is the nucleon mass,  $\rho_{\rm sour}$  is the hyperradius,  $\hbar$  is the reduced Planck constant and the factor 3.3 was estimated by matching to the calculated distribution in ref. <sup>13</sup>.

## Background events from competing reactions

Events with a proton and an  $\alpha$ -particle in the final state might result from other competing reactions. However, owing to the unique kinematics of quasi-elastic p- ${}^4\text{He}$  scattering at large centre-of-mass angles, other direct reactions are excluded. Examples are one-neutron knockout,  $\{(\{\})^8\{\rm He\}\}(\{\{\rm p\}\}, \{\{\rm p\}\}\{\{\rm n\}\})^7\{\rm He\}$ , or  ${}^6\text{He}$  knockout,  $\{(\{\})^8\{\rm He\}\}(\{\{\rm p\}\}, \{\{\rm p\}\})^6\{\{\rm He\}\}^*$ . In the former case, the momentum of the proton is too small to be selected by the experiment, while in the latter, it is too high at the angular range covered by the set-up, owing to the larger mass of  ${}^6\text{He}$  relative to  ${}^4\text{He}$ . Therefore, the only possible contribution of background events to our measured p- ${}^4\text{He}$  events can come from secondary reactions. Below we list the possible reactions and their expected contributions.

$\langle\langle\}^8\{\rm{He}\}\{\{\rm{p}\},\{\rm{pn}\}\}\rangle\rangle^7\{\rm{He}\}\rangle\rangle$ . A first interaction is single-neutron knockout from  ${}^8\text{He}$  leading to break-up into  ${}^6\text{He}$  and two neutrons. This can be followed by a second interaction, quasi-elastic p– ${}^4\text{He}$  scattering at backward centre-of-mass angles. These processes will result in events with a missing-mass distribution similar to that of the one-step  $\langle\langle\}^6\{\rm{He}\}\{\{\rm{p}\},\{\{\rm{p}\}\}^4\{\rm{He}\}\rangle\rangle$  reaction but broadened and shifted to smaller energies. This is due to the difference in the separation energies of two and four neutrons in  ${}^6\text{He}$  versus  ${}^8\text{He}$  of about 2 MeV, and due to the additional recoil ([Supplementary Information](#)). The single-neutron knockout cross-sections from  ${}^{6,8}\text{He}$  were measured in an experiment performed within the same experimental campaign as the experiment presented here, at a slightly higher beam energy of 185 MeV per nucleon using a series of targets with different nuclear charges  $Z$  (ref. [30](#)). For  ${}^6\text{He}$ , the cross-section was extracted for the hydrogen target, whereas for  ${}^8\text{He}$ , the lowest- $Z$  target used was carbon. We therefore scale the measured cross-section for  ${}^6\text{He}$  with hydrogen,  $\langle\langle\sigma\}_{\langle\langle}^6\{\rm{He}\},\{\rm{H}\}\rangle\rangle$ , by the carbon measurement  $\langle\langle\sigma\}_{\langle\langle}^8\{\rm{He}\},\{\rm{C}\}\rangle\rangle/\langle\langle\sigma\}_{\langle\langle}^6\{\rm{He}\},\{\rm{C}\}\rangle\rangle=1.26\pm 0.16$ . Using  $\langle\langle\sigma\}_{\langle\langle}^6\{\rm{He}\},\{\rm{H}\}\rangle\rangle=47\pm 4\,\{\rm{mb}\}$  results in  $\langle\langle\sigma\}_{\langle\langle}1\{\rm{n}\}\rangle\rangle=59\pm 9\,\{\rm{mb}\}$ .

$\langle\langle\}^8\{\rm{He}\}\{\{\rm{p}\},\{\rm{p}\}\}\rangle\rangle^8\{\rm{He}\}\rangle\rangle^{*\rangle\rangle}$ . A first interaction is inelastic excitation of  ${}^8\text{He}$  leading to a break-up into  ${}^6\text{He}$  and two neutrons. These cross-sections were measured as well in ref. [30](#), with scaling ratio of  $\langle\langle\sigma\}_{\langle\langle}^8\{\rm{He}\},\{\rm{C}\}\rangle\rangle/\langle\langle\sigma\}_{\langle\langle}^6\{\rm{He}\},\{\rm{C}\}\rangle\rangle=1.02\pm 0.13$ . Using  $\langle\langle\sigma\}_{\langle\langle}^6\{\rm{He}\},\{\rm{H}\}\rangle\rangle=11.0\pm 0.9\,\{\rm{mb}\}$  gives  $\langle\langle\sigma\}_{\langle\langle}\{\rm{inel}\}\rangle\rangle=11.2\pm 1.7\,\{\rm{mb}\}$ .

In total, the cross-section evaluated for the two processes described above equals  $\langle\langle\sigma\}_{\langle\langle}\{\rm{tot}\}\rangle\rangle=71.1\pm 9.2\,\{\rm{mb}\}$ . Evaluating the relative contribution to the number of one-step p– ${}^4\text{He}$  quasi-elastic reactions  $N_{\rm reac}$  at half of the target thickness ( $t$ ) gives

$$\begin{aligned}
 & (\langle N \rangle_{\langle\langle} \{\rm{bkg}\} \rangle / \langle N \rangle_{\langle\langle} \{\rm{reac}\} \rangle) = \sigma(\{\rm{b}\}) \times t(\{\rm{g}\}) \times (\{\rm{cm}\})^{-2} \times \\
 & (0.6(\{\rm{Avogadro}\}) / \{\rm{A}\}) = 0.071 \times 0.175 \times 0.6 \\
 & (= 0.75 \pm 0.10 \%), \text{ where } A \text{ is the mass number of the target. These}
 \end{aligned}$$

types of processes can also occur from  $^8\text{He}$  break-up along beamline materials before the target, plastic scintillators, kapton windows and air. These contributions were estimated starting from the last plastic scintillators at the BigRIPS fragment separator up to the kapton window at the entrance to the target cell, using the measured cross-sections for hydrogen and carbon targets<sup>30</sup>. The produced  $^6\text{He}$  has an angular spread in the transverse direction according to its intrinsic momentum, which has a width of  $\langle \sigma \rangle_{\rm He} = 35 \text{ MeV}/c$  (ref. 18). Assuming a central beam energy of 156 AMeV, its corresponding momentum is  $\langle P \rangle_{\rm He} = 3.37 \text{ GeV}/c$ , leading to an angular  $\theta$  width of  $\langle \sigma_\theta \rangle = \langle \sigma \rangle_{\rm He} / \langle P \rangle_{\rm He} = 0.035/3.37 = 10 \text{ mrad}$ . The spread in the  $x$  and  $y$  directions evaluated at the entrance to the target equals  $\langle \sigma_{x,y} \rangle = d \cdot \langle \sigma_\theta \rangle = d \text{ mm} \cdot 10 \text{ mrad}$ , where  $d$  is the distance from the material at which  $^6\text{He}$  was produced to the target entrance. The target enclosure radius equals 20 mm. Therefore, depending on the distance of production point and target, some fraction of the produced  $^6\text{He}$  is expected to not reach the target, which reduces its intensity. Using the measured distances to the target entrance, the estimated fraction of  $^6\text{He}$  produced before the target and reaching the effective target volume is  $0.36 \pm 0.01\%$ . Overall, the estimated contribution from these two processes is  $1.11 \pm 0.10\%$ .

$\langle \sigma \rangle_{\rm He} = 35 \text{ MeV}/c$ . A first interaction is  $^6\text{He}$  knockout to its ground state. This process can contribute only when the  $^6\text{He}$  is produced along the target, as, owing to its angular spread, it will not reach the target region in case of production in beamline materials. The cross-section for this process is not well known, and was studied previously only for a much higher energy<sup>18</sup>. However, it was shown to be compatible with the p– $^4\text{He}$  elastic scattering cross-section. We therefore adapt the measured p– $^4\text{He}$  total cross-section, which was measured at 156 MeV (ref. 16),  $\sigma = 91.8 \text{ mb}$ , and use it as an upper limit. The relative contribution of these events is evaluated as  $(0.0918 \times 0.175 \times 0.6 = 0.96\%)$ .

$\langle \sigma \rangle_{\rm He} + \langle \sigma \rangle_{\rm p} \cdot \langle \sigma \rangle_{\rm He}$ . In the first interaction,  $^4\text{He}$  is produced from fragmentation of  $^8\text{He}$ . This can be, for

example, a result of elastic p– ${}^4\text{He}$  scattering or inelastic excitation of  ${}^8\text{He}$ . The inclusive  ${}^4\text{He}$  cross-section was measured at higher energy for  $\langle\{\}^8\{\text{rm}\{\text{He}\}\}+\{\}^{12}\{\text{rm}\{\text{C}\}\}\rangle$  to  $\langle\{\}^4\{\text{rm}\{\text{He}\}\}\rangle$  to be  $95 \pm 5 \text{ mb}$  (ref. [31](#)). We scale it by a factor of 1/2 to estimate the cross-section for the hydrogen target, leading to  $\sigma = 48 \pm 5 \text{ mb}$ . The relative contribution of these events is evaluated as  $\langle 0.048 \times 0.175 \times 0.6 = 0.50 \text{ pm } 0.05 \%$ . This process will result in events with a negative missing-mass distribution below  $-3.1 \text{ MeV}$ , the binding energy of  ${}^4\text{He}$  in  ${}^8\text{He}$  that will extend to more negative values owing to the additional recoil ([Supplementary Information](#)).

$\langle\{\}^8\{\text{rm}\{\text{He}\}\}\langle\{\text{rm}\{\text{p}\}\},\{\text{rm}\{\text{p}\}\}\rangle\rangle^8\{\text{rm}\{\text{He}\}\}$ . A first interaction is p– ${}^8\text{He}$  elastic scattering at backwards centre-of-mass angles. This will lead to a fast proton in the final state. In a second interaction,  ${}^4\text{He}$  can be produced. However, in this case, the momentum of the proton is even larger than that resulting from direct  ${}^6\text{He}$  knockout, owing to the larger mass; therefore, this reaction channel is excluded.

For the benchmark measurement with  ${}^6\text{He}$ , only one reaction can contribute to a background, where  ${}^4\text{He}$  is produced in the first interaction,  $\langle\{\}^6\{\text{rm}\{\text{He}\}\}+\{\text{rm}\{\text{p}\}\}\rangle$  to  $\langle\{\}^4\{\text{rm}\{\text{He}\}\}\rangle$ . Similar to  ${}^8\text{He}$ , the inclusive  ${}^4\text{He}$  cross-section was measured at higher energy for  $\langle\{\}^6\{\text{rm}\{\text{He}\}\}+\{\}^{12}\{\text{rm}\{\text{C}\}\}\rangle$  to  $\langle\{\}^4\{\text{rm}\{\text{He}\}\}\rangle$  to be  $189 \pm 14 \text{ mb}$  (ref. [31](#)). We scale it by a factor of 1/2 to estimate the cross-section for the hydrogen target, leading to  $\sigma = 95 \pm 7 \text{ mb}$ . The relative contribution of these events is evaluated as  $\langle 0.095 \times 0.175 \times 0.6 = 1.00 \text{ pm } 0.07 \%$ . This process will result in events with a negative missing-mass distribution below  $-0.975 \text{ MeV}$ , the binding energy of  ${}^4\text{He}$  in  ${}^6\text{He}$  that will extend to more negative values owing to the additional recoil ([Supplementary Information](#)).

## Summary of background contributions

From the reactions listed above, we conclude that the only contributions from background processes come from reactions in which  ${}^6\text{He}$  or  ${}^4\text{He}$  were produced in a first step, and the overall contribution is evaluated as 2.6%. In the fitting of the missing-mass spectrum,  $f_{\text{bgk}}$  is taken as the simulated distribution for the two-step processes described above. Each process is

simulated individually, and adds weight to the total distribution according to its expected contribution. In addition, as the measured cross-sections adopted to estimate the background contributions were not measured directly for the reactions of interest, we consider the uncertainty as a variation by a factor of two on each one of the scaled cross-sections (Supplementary Information).

It is noted that the resulting background distribution can also explain the two events observed at the energy region of  $\langle -10, \text{MeV} \rangle < \{E\}_{4\{\text{rm}\{n\}\}} < -1.8, \{\text{MeV}\}.$ ) These two events cannot result from a direct reaction, as the lower limit of a bound tetraneutron is  $-3.1$  MeV, corresponding to the binding energy of  ${}^8\text{He}$  against decay into  $(\{\}^4\{\text{rm}\{\text{He}\}\} + 4\{\text{rm}\{n\}\})$ . Therefore, these events can be attributed only to two-step reactions. Estimating the background function at  $-9$  MeV, we expect  $0.27$  events in that energy region, leading to a  $2\sigma$  significance of the measured two events, which is not considered as a statistically significant deviation from the expected background contribution in that energy region.

Finally, a major difference between one-step and two-step processes is expected in the reaction-vertex distribution of the proton and the  $\alpha$ -particle (Extended Data Fig. 8). For a two-step process, an exponentially increasing yield of reaction products along the target is expected owing to the need to produce the  ${}^6\text{He}$  ( ${}^4\text{He}$ ) first. Such a distribution is not observed in the data, corroborating that the background contribution is indeed small. The estimated background contribution will be considered to evaluate the significance level of the observed peak.

To quantify the statistical significance of the resonance observed in Fig. 3, we evaluate the number of background events in the peak region defined as  $\langle E_{\{\{\text{rm}\{r\}\}\}} \pm 2\text{varGamma} \rangle$  and  $\langle -2, \{\text{MeV}\} \rangle < \{E\}_{4\{\text{rm}\{n\}\}} < 6, \{\text{MeV}\}.$ ) The number of measured events in the peak region amounts to  $N = 41$  events. The number of background events equals the integral of the background and continuum functions ( $f_{\text{bkg}}$  and  $f_{\text{cont}}$  in equation (2)) in that region, resulting in an average number of eight events from the background and two events from the continuum, such that the total expected background amounts to ten events. Even though we allowed a

variation of a factor of two for the background estimation, this does not change the result significantly. Hence, we conclude that the resonance structure is observed with high statistical significance, well above the  $5\sigma$  observation threshold.

## Data availability

The datasets generated during and/or analysed during the current study are available from the corresponding author on reasonable request. [Source data](#) are provided with this paper.

## References

1. Marqués, F. M. & Carbonell, J. The quest for light multineutron systems. *Eur. Phys. J. A* **57**, 105 (2021).
2. Marqués, F. M. et al. Detection of neutron clusters. *Phys. Rev. C* **65**, 044006 (2002).
3. Marqués, F. M., Orr, N. A., Al Falou, H., Normand, G. & Clarke, N. M. On the possible detection of  ${}^4n$  events in the breakup of  ${}^{14}\text{Be}$ . Preprint at <https://arxiv.org/abs/nucl-ex/0504009> (2005).
4. Kisamori, K. et al. Candidate resonant tetraneutron state populated by the  ${}^4\text{He}({}^8\text{He}, {}^8\text{Be})$  reaction. *Phys. Rev. Lett.* **116**, 052501 (2016).
5. Gandolfi, S., Hammer, H.-W., Klos, P., Lynn, J. E. & Schwenk, A. Is a trineutron resonance lower in energy than a tetraneutron resonance? *Phys. Rev. Lett.* **118**, 232501 (2017).
6. Schiffer, J. P. & Vanderbosch, R. Search for a particle-stable tetra neutron. *Phys. Lett.* **5**, 292–293 (1963).
7. Unger, J. E. et al. Search for the tetraneutron by the double-charge-exchange of negative pions. *Phys. Lett. B* **144**, 333–336 (1984).

8. Beaumel, D. et al. Studies of exotic systems using transfer reactions at GANIL. In *Proc. International Symposium on the ISPUN07* 8–25 (World Scientific, 2008).
9. Bertulani, C. A. & Zelevinsky, V. Is the tetraneutron a bound dineutron–dineutron molecule? *J. Phys. G* **29**, 2431–2437 (2003).
10. Pieper, S. C. Can modern nuclear Hamiltonians tolerate a bound tetraneutron? *Phys. Rev. Lett.* **90**, 252501 (2003).
11. Timofeyuk, N. K. Do multineutrons exist? *J. Phys. G* **29**, L9–L14 (2003).
12. Zhukov, M. V., Korsheninnikov, A. A. & Smedberg, M. H. Simplified  $\alpha+4n$  model for the  ${}^8\text{He}$  nucleus. *Phys. Rev. C* **50**, R1–R4 (1994).
13. Grigorenko, L. V., Timofeyuk, N. K. & Zhukov, M. V. Broad states beyond the neutron drip line. *Eur. Phys. J. A* **19**, 187–201 (2004).
14. Kobayashi, T. et al. SAMURAI spectrometer for RI beam experiments. *Nucl. Instrum. Methods B* **317**, 294–304 (2013).
15. Obertelli, A. et al. MINOS: a vertex tracker coupled to a thick liquid-hydrogen target for in-beam spectroscopy of exotic nuclei. *Eur. Phys. J. A* **50**, 8 (2014).
16. Comparat, V. et al. Elastic proton scattering on  ${}^4\text{He}$  at 156 MeV. *Phys. Rev. C* **12**, 251–255 (1975).
17. Göbel, M. et al. Neutron–neutron scattering length from the  ${}^6\text{He}(\text{p},\text{p}\alpha)\text{nn}$  reaction. *Phys. Rev. C* **104**, 024001 (2021).
18. Chulkov, L. V. et al. Quasi-free scattering with  ${}^{6,8}\text{He}$  beams. *Nucl. Phys. A* **759**, 43–63 (2005).
19. Shirokov, A. M. et al. Prediction for a four-neutron resonance. *Phys. Rev. Lett.* **117**, 182502 (2016).

20. Li, J. G., Michel, N., Hu, B. S., Zuo, W. & Xu, F. R. Ab-initio no-core Gamow shell-model calculations of multineutron systems. *Phys. Rev. C* **100**, 054313 (2019).
21. Fossez, K., Rotureau, J., Michel, N. & Płoszajczak, M. Can tetraneutron be a narrow resonance? *Phys. Rev. Lett.* **119**, 032501 (2017).
22. Sofianos, S. A., Rakityansky, S. A. & Vermaak, G. P. Subthreshold resonances in few-neutron systems. *J. Phys. G* **23**, 1619–1629 (1997).
23. Lazauskas, R. & Carbonell, J. Is a physically observable tetraneutron resonance compatible with realistic nuclear interactions? *Phys. Rev. C* **72**, 034003 (2005).
24. Hiyama, E., Lazauskas, R., Carbonell, J. & Kamimura, M. Possibility of generating 4-neutron resonance with  $T=3/2$  isospin 3-neutron force. *Phys. Rev. C* **93**, 044004 (2016).
25. Lazauskas, R., Carbonell, J. & Hiyama, E. Modeling the double charge exchange response function for a tetraneutron system. *Prog. Theor. Exp. Phys.* **2017**, 073D03 (2017).
26. Deltuva, A. Tetraneutron: rigorous continuum calculation. *Phys. Lett. B* **782**, 238–241 (2018).
27. Higgins, M. D., Greene, C. H., Kievsky, A. & Viviani, M. Nonresonant density of states enhancement at low energies for three or four neutrons. *Phys. Rev. Lett.* **125**, 052501 (2020).
28. Thompson, I. J., Nunes, F. M. & Danilin, B. V. FaCE: a tool for three body Faddeev calculations with core excitation. *Comp. Phys. Commun.* **161**, 87–107 (2004).
29. Timofeyuk, N. K. Shell model approach to construction of a hyperspherical basis for  $A$  identical particles: application to hydrogen and helium isotopes. *Phys. Rev. C* **65**, 064306 (2002).

30. Lehr, C. *Low-energy Dipole Response of the Halo Nuclei  $^{6,8}He$* . PhD thesis, TU Darmstadt (2022); <https://tuprints.ulb.tu-darmstadt.de/20267/>
31. Tanihata, I. et al. Revelation of thick neutron skins in nuclei. *Phys. Lett. B* **289**, 261–266 (1992).

## Acknowledgements

We thank M. Göbel and H. W. Hammer for providing the theoretical input of the two-neutron energy distribution. This work was supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) Project-ID 279384907 - SFB 1245, the GSI-TU Darmstadt cooperation agreement, by the UK STFC under contract numbers ST/P003885/1 and ST/L005727/1 and the University of York Pump Priming Fund, by the German Federal Ministry of Education and Research - BMBF project numbers 05P15RDFN1, 05P15WOFNA and 05P15WOCIA, by project FAIR-RO-04/DEMAND - IFA, by JSPS KAKENHI grant numbers JP16H02177, JP16H02179 and JP18H05404, by the Spanish Research grant PGC2018-099746-B-C21, and by the Swedish Research Council, project grant 2011-5324 and 2017-03839. K.I.H., D. Kim, S.K., L.S., S.Y.P. and D.A. acknowledge the support from the IBS grant funded by the Korea government grant number IBS-R031-D1. C.A.B. acknowledges partial support by the US DOE grant number DE-FG02- 08ER41533. I.G. acknowledges support by HIC for FAIR and Croatian Science Foundation under project numbers 1257 and 7194. Z.E., Z.H. and I.K. acknowledge the support by NKFIH grants numbers 114454, 128947 and GINOP-2.3.3-15-2016-00034. The LPC group acknowledge support from the France-Japanese LIA-International Associated Laboratory for Nuclear Structure Problems as well as the French ANR-14-CE33-0022-02 EXPAND.

## Author information

### Authors and Affiliations

1. Technische Universität Darmstadt, Fachbereich Physik, Darmstadt, Germany  
M. Duer, T. Aumann, S. Paschalis, D. M. Rossi, C. Caesar, I. Gasparic, M. Holl, J. Kahlbow, S. Kim, P. Koseoglou, C. Lehr, K. Miki, A. Obertelli, H. Scheit, F. Schindler, D. Symochko, J. Tanaka, H. T. Törnqvist, J. Tscheuschner, V. Wagner & L. Zanetti
2. GSI Helmholtzzentrum für Schwerionenforschung GmbH, Darmstadt, Germany  
T. Aumann, V. Panin, K. Boretzky, C. Caesar, D. Körper, H. Simon & H. T. Törnqvist
3. Helmholtz Forschungsakademie Hessen für FAIR, Frankfurt, Germany  
T. Aumann
4. Technische Universität München, Physik Department, Garching, Germany  
R. Gernhäuser, M. Böhmer, F. Dufter & S. Reichert
5. RIKEN Nishina Center for Accelerator-Based Science, Wako, Japan  
V. Panin, D. Ahn, H. Baba, C. Caesar, N. Chiga, N. Fukuda, I. Gasparic, Z. Ge, N. Inabe, T. Isobe, Y. Kubota, H. Otsu, M. Sasano, H. Suzuki, H. Takeda, J. Tanaka, T. Uesaka, Z. H. Yang, K. Yoneda & J. Zenihiro
6. Department of Physics, University of York, York, UK  
S. Paschalis & U. Forsberg
7. LPC-Caen, IN2P3/CNRS, UniCaen, ENSICAEN, Normandie Université, Caen, France  
N. L. Achouri, J. Gibelin, F. M. Marqués, B. Monteagudo, N. A. Orr, M. Parlog & A. Revel

8. Department of Physics and Astronomy, Texas A&M University-Commerce, Commerce, TX, USA  
C. A. Bertulani
9. IRFU, CEA, Université Paris-Saclay, Gif-sur-Yvette, France  
A. Corsi, J. M. Gheller, A. Gillibert, A. Obertelli & A. Revel
10. Dpt. de Física de Partículas, Universidade de Santiago de Compostela, Santiago de Compostela, Spain  
D. Cortina-Gil & B. Fernández-Domínguez
11. Nuclear Energy group, ESRIG, University of Groningen, Groningen, The Netherlands  
C. A. Douma, M. N. Harakeh & N. Kalantar-Nayestanaki
12. Atomki, Eötvös Loránd Research Network (ELKH), Debrecen, Hungary  
Z. Elekes, Z. Halász & I. Kuti
13. State Key Laboratory of Nuclear Physics and Technology, School of Physics, Peking University, Beijing, China  
J. Feng, Y. Liu & B. Yang
14. Rudjer Bošković Institute, Zagreb, Croatia  
I. Gasparic
15. Ewha Womans University, Seoul, Korea  
K. I. Hahn & S. Y. Park
16. Center for Exotic Nuclear Studies, Institute for Basic Science, Daejeon, Korea

D. Ahn, K. I. Hahn, D. Kim, S. Kim, S. Y. Park & L. Stuhl

17. Department of Physics, Tokyo Institute of Technology, Tokyo, Japan

A. Hirayama, Y. Kondo, M. Matsumoto, T. Nakamura, A. T. Saito, Y. Togano, T. Tomai, H. Yamada & M. Yasuda

18. Department of Physics, Tohoku University, Miyagi, Japan

T. Kobayashi & K. Miki

19. Department of Physics, The University of Hong Kong, Hong Kong, China

P. J. Li

20. Department of Physics, Chalmers University of Technology, Gothenburg, Sweden

S. Lindberg, T. Nilsson & M. V. Zhukov

21. Center for Nuclear Study, The University of Tokyo, Tokyo, Japan

S. Masuoka, S. Shimoura, L. Stuhl & L. Yang

22. Universität zu Köln, Institut für Kernphysik, Cologne, Germany

J. Mayer

23. Institute of Space Sciences, Magurele, Romania

P. M. Potlog

24. GANIL, CEA/DRF-CNRS/IN2P3, Caen, France

A. Revel

25. Department of Physics, Kyoto University, Kyoto, Japan

J. Zenihiro

## Contributions

The experimental set-up at the RIKEN Nishina Center was designed and constructed by the SAMURAI Collaboration. The design of the present experiment was performed by a large number of SAMURAI Collaboration members, led by T.A., S.P., D.M.R. and V.P. In particular, design and construction of the silicon trackers was lead by R.G. The main data analysis was undertaken by M.D. All authors contributed to setting up and data taking, and all authors read and approved the manuscript.

## Corresponding author

Correspondence to [M. Duer](#).

## Ethics declarations

### Competing interests

The authors declare no competing interests.

## Peer review

### Peer review information

*Nature* thanks Stefano Gandolfi, Michele Viviani and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

## Additional information

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

## Extended data figures and tables

## Extended Data Fig. 1 Incoming beam identification and tracking.

(a) Charge identification vs. time-of-flight (TOF) for incoming beam ions measured event-by-event using plastic scintillators. The TOF is measured relative to the last scintillators at the BigRIPS fragment separator. The red ellipse represents the cut used to identify  ${}^8\text{He}$  ions. The peak above  ${}^8\text{He}$ , with the same TOF is attributed to pile-up events. The peak at smaller TOF and  $Z=3$  originates from Li. (b) Angular profile of the incoming beam in the XY plane measured by two drift chambers

[Source data](#)

## Extended Data Fig. 2 Reaction vertex.

(a) 3D representation of the reconstructed reaction vertex from empty target runs. In such a case the incoming beam particle can interact with the material of the target cell at the entrance/exit of the cell, or with the residual hydrogen gas inside the target cell. (b) Reconstructed vertex in the XY plane. The region outside the solid circle is excluded in the analysis, to avoid reactions produced at the target cell and the support structure of the target. The dashed circle indicates the target cell, which has radius of 20 mm. (c) Projection onto the  $z$ -axis. The red lines represent the cut applied in the analysis. By fitting the peaks with Gaussians distributions, the vertex resolution is extracted to be  $\langle \sigma \rangle = 1.69 \pm 0.04 \, \text{mm}$  at the entrance of the target, and  $\langle \sigma \rangle = 0.71 \pm 0.01 \, \text{mm}$  at the exit of the target. The first silicon plane in these plots is located at  $z = 0 \, \text{mm}$

[Source data](#)

## Extended Data Fig. 3 Target reconstruction.

Reconstructed reaction vertex with the liquid-hydrogen target in ZX (a), ZY (b), and XY (c) planes. The cylindrical shape of the target is visible with a thickness of  $\sim 50 \, \text{mm}$ , where the first silicon plane in these plots is located at  $z = 0$ . In  $x$ - and  $y$ -directions, the vertices are Gaussian distributed with a

cutoff according to the target cell radius. In y-direction a lower upper cutoff is visible, indicating that the target cell was not completely filled by liquid-hydrogen

[Source data](#)

### **Extended Data Fig. 4 Charged fragments identification.**

Aligned energy deposition at each of the scintillator walls vs. the measured position in the drift chamber for events with two fragments. Different bands correspond to different fragments which are labelled. The left (right) scintillator wall (viewing opposite to the beam direction in Fig. 2) covers fragments emerging with relatively small (large) momenta. The red ellipses represent the cuts used to identify ‘slow’  ${}^4\text{He}$  and ‘fast’ protons (see [Supplementary Information](#))

[Source data](#)

### **Extended Data Fig. 5 Energy response of the set-up.**

(a) Generated vs. reconstructed energies of the four-neutron system for a uniformly distributed generated energy in the range  $\langle -60 < \{\text{rm}\{E\}\} \rangle_4 \{\boldsymbol{n}\}, \{\text{rm}\{\text{gen}\}\} \rangle < 100, \{\text{rm}\{\text{MeV}\}\}, \rangle$  transported through the experimental set-up. A clear diagonal correlation between the generated and reconstructed energies is observed. The colour code represents the acceptance of the set-up, normalized such that the maximum corresponds to one. (b) Projection of the acceptance as a function of the generated energy. (c) Projection of the response, as an example, for generated energy of zero MeV. The fitted width of the distribution of 1 MeV corresponds to the resolution of the reconstructed energy

[Source data](#)

### **Extended Data Fig. 6 Energy of the 2n system.**

Theoretical  $E_{2n}$  distribution after  ${}^4\text{He}$  knockout from  ${}^6\text{He}$  resulting from the  ${}^6\text{He}$  ground-state wavefunction with (purple) and without (green) taking into

account  $nn$  final-state interaction (FSI). Distributions are normalized such that their maximum is the same

[Source data](#)

### **Extended Data Fig. 7 Estimation of the systematic uncertainty.**

(a) Calculated  $\chi^2$  values as a function of the offset added to the theoretical distribution used to describe the energy of the two-neutron system in the  $\langle \{ \}^6 \{ \rm H \} \{ \rm e \} (\{ \rm p \}, \{ \rm p \}) \rangle^4 \{ \rm H \} \{ \rm e \} \rangle$  reaction. The systematic uncertainty for the energy is taken as the offset values at  $\langle \{ \chi \}_\{\{ \rm min \} \} \rangle^2 \pm 1,$  shown by the dashed lines. (b) The corresponding distributions of the data (black symbols), the original (solid red) and the shifted (dashed red) theoretical curves. Distributions are presented for the low-energy region, where the difference between the theory curves is visible. (c) Same as (a) but as a function of the energy resolution (using an offset of zero). The systematic uncertainty for the width is taken as the resolution values at  $\langle \{ \chi \}_\{\{ \rm min \} \} \rangle^2 \pm 1,$  shown by the dashed lines. (d) The corresponding distributions of the data (black symbols), the original (solid red) and the  $\langle \{ \chi \}_\{\{ \rm min \} \} \rangle^2 \pm 1$  resolution smeared (dashed red) theoretical curves. For comparison, the original distribution is also shown, where only the acceptance of the set-up has been taken into account but not the resolution, *i.e.*, a resolution of zero MeV has been assumed. The distribution is scaled by a factor of 1/2 to match the scale of the figure

[Source data](#)

### **Extended Data Fig. 8 Reaction vertex of the selected events.**

Reaction vertex for the selected quasi-elastic  $\langle \{ \}^8 \{ \rm H \} \{ \rm e \} \rangle (\{ \rm p \}, \{ \rm p \}) \rangle^4 \{ \rm H \} \{ \rm e \} \rangle$  events. Data sample is shown by the full symbols, and the low-energy peak region is shown by the open symbols and normalized according to the number of events for the full data set, together with a simulation of one-step process (red curve), and two-step process (blue curve): break-up of  ${}^8\text{He}$  into  ${}^6\text{He}$  followed by quasi-elastic  $\langle \{ \rm p \} - \{ \}^4 \{ \rm H \} \{ \rm e \} \rangle$  scattering. As can be seen, an

exponential increase is expected towards the exit of the target, which is not observed in the data, confirming that the contribution from this background is indeed as small as expected from the evaluation based on measured cross-sections. Simulated distributions for other two-step processes show the same exponential behaviour

## [Source data](#)

# **Supplementary information**

## [Supplementary Information](#)

This Supplementary Information file contains the following sections. Section 1, SAMURAI set-up; Section 2, Beam measurement; Section 3, Fragments measurement; Section 4, Quasi-elastic events. It includes Supplementary Figs, 1–29, Tables 1–4 and references.

## **Source data**

### [Source Data Fig. 2](#)

### [Source Data Fig. 3](#)

### [Source Data Fig. 4](#)

### [Source Data Extended Data Fig. 1](#)

### [Source Data Extended Data Fig. 2](#)

### [Source Data Extended Data Fig. 3](#)

### [Source Data Extended Data Fig. 4](#)

### [Source Data Extended Data Fig. 5](#)

## [Source Data Extended Data Fig. 6](#)

## [Source Data Extended Data Fig. 7](#)

## [Source Data Extended Data Fig. 8](#)

## Rights and permissions

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

## Reprints and Permissions

## About this article

### Cite this article

Duer, M., Aumann, T., Gernhäuser, R. *et al.* Observation of a correlated free four-neutron system. *Nature* **606**, 678–682 (2022).  
<https://doi.org/10.1038/s41586-022-04827-6>

- Received: 04 August 2021
- Accepted: 28 April 2022
- Published: 22 June 2022

- Issue Date: 23 June 2022
- DOI: <https://doi.org/10.1038/s41586-022-04827-6>

## Share this article

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

[Get shareable link](#)

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

[Copy to clipboard](#)

Provided by the Springer Nature SharedIt content-sharing initiative

## [\*\*Collisions hint that four neutrons form a transient isolated entity\*\*](#)

- Lee G. Sobotka
- Maria Piarulli

Nature News & Views 22 Jun 2022

---

This article was downloaded by **calibre** from <https://www.nature.com/articles/s41586-022-04827-6>

- Article
- Open Access
- [Published: 08 June 2022](#)

# Continuous Bose–Einstein condensation

- [Chun-Chia Chen](#),
- [Rodrigo González Escudero](#),
- [Jiří Minář](#),
- [Benjamin Pasquiou](#),
- [Shayne Bennetts](#) &
- [Florian Schreck](#)

*Nature* volume **606**, pages 683–687 (2022)

- 15k Accesses
- 611 Altmetric
- [Metrics details](#)

## Abstract

Bose–Einstein condensates (BECs) are macroscopic coherent matter waves that have revolutionized quantum science and atomic physics. They are important to quantum simulation<sup>1</sup> and sensing<sup>2,3</sup>, for example, underlying atom interferometers in space<sup>4</sup> and ambitious tests of Einstein’s equivalence principle<sup>5,6</sup>. A long-standing constraint for quantum gas devices has been the need to execute cooling stages time-sequentially, restricting these devices to pulsed operation. Here we demonstrate continuous Bose–Einstein condensation by creating a continuous-wave (CW) condensate of strontium atoms that lasts indefinitely. The coherent matter wave is sustained by

amplification through Bose-stimulated gain of atoms from a thermal bath. By steadily replenishing this bath while achieving 1,000 times higher phase-space densities than previous works<sup>7,8</sup>, we maintain the conditions for condensation. Our experiment is the matter wave analogue of a CW optical laser with fully reflective cavity mirrors. This proof-of-principle demonstration provides a new, hitherto missing piece of atom optics, enabling the construction of continuous coherent-matter-wave devices.

## Main

Continuous operation is advantageous for sensors as it eliminates dead time and can offer higher bandwidths than pulsed operation<sup>9,10,11,12</sup>. Meanwhile, sensors using BECs benefit from their high phase-space density and unique coherence properties<sup>2,3,4,5,6,13</sup>. Combining these advantages, a CW atom laser beam outcoupled from a CW condensate could be ideal for many quantum sensing applications<sup>14,15,16,17</sup>. In the long term, CW atom lasers could benefit applications ranging from dark-matter and dark-energy searches<sup>18,19</sup>, gravitational-wave detection<sup>20,21,22,23,24</sup>, tests of Einstein's equivalence principle<sup>5,6</sup> to explorations in geodesy<sup>25,26,27</sup>. In the short term, the CW BEC offers a platform to study quantum atom optics and new quantum phenomena arising in driven-dissipative quantum gases<sup>28</sup>.

The key to realizing a CW BEC of atoms is to continuously amplify the atomic matter wave while preserving its phase coherence<sup>29</sup>. An amplification process is essential to compensate naturally occurring atom losses, for example, from molecule formation. It is also needed to replace the atoms that will be coupled out of the BEC for sustaining an atom laser or atom interferometer. Addressing this challenge requires two ingredients: a gain mechanism that amplifies the BEC and a continuous supply of ultracold atoms near quantum degeneracy.

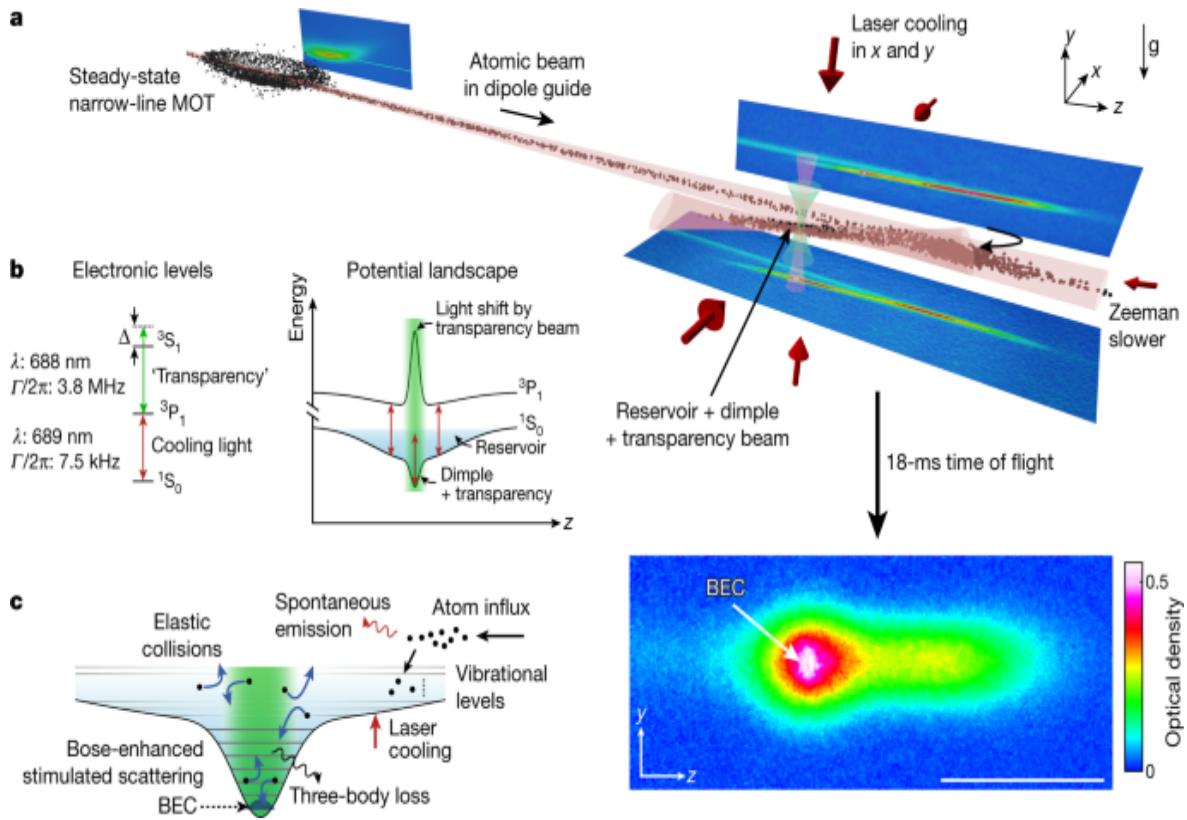
The first steps towards a continuous gain mechanism were taken in ref. <sup>30</sup>, in which merging of independent condensates periodically added atoms to an already existing BEC, but in which coherence was not retained across merger events. A Bose-stimulated gain mechanism into a single dominant mode (the BEC) is required to provide gain without sacrificing phase

coherence. Such gain mechanisms have been demonstrated using elastic collisions between thermal atoms<sup>29,31</sup>, stimulated photon emission<sup>32</sup>, four-wave mixing<sup>33,34</sup> or superradiance<sup>35</sup>. However, in all these demonstrations, the gain mechanism could not be sustained indefinitely.

To sustain gain, the second ingredient is needed: a continuous supply of ultracold, dense gas with a phase-space density—the occupancy of the lowest motional quantum state—approaching  $\rho = 1$ . Great efforts were spent developing continuously cooled beams of atoms<sup>36,37,38,39</sup> and continuously loaded traps<sup>7,8</sup>, which—so far—have reached phase-space densities of  $\rho = 10^{-3}$ . To achieve the required microkelvin temperatures, these experiments have to use laser cooling, but near-resonant laser-cooling light is highly destructive for BECs<sup>40</sup>. Several experiments have maintained a BEC in the presence of harmful light, either by spatially separating the laser cooling from the quantum gas<sup>7,30,36,39,41</sup> or by reducing the absorbance of the quantum gas<sup>29,42,43,44</sup>.

Here we demonstrate the creation of a CW BEC that can last indefinitely. Our experiment comprises both ingredients, gain and continuous supply, as illustrated in Fig. 1. The centrepiece of the experiment consists of a large ‘reservoir’ that is continuously loaded with Sr atoms and that contains a small and deep ‘dimple’ trap in which the BEC is created. The gas in the reservoir is continuously laser-cooled and exchanges atoms and heat with the dimple gas. A ‘transparency’ beam renders atoms in the dimple transparent to harmful laser-cooling photons. The dimple increases the gas density while the temperature is kept low by thermal contact with the reservoir. This enhances the phase-space density, leading to the formation of a BEC. Bose-stimulated elastic collisions continuously scatter atoms into the BEC mode, providing the gain necessary to sustain it indefinitely.

**Fig. 1: Experimental setup and scheme.**



**a**,  $^{84}\text{Sr}$  atoms from a steady-state narrow-line magneto-optical trap (MOT) are continuously outcoupled into a guide and loaded into a crossed-beam dipole trap that forms a large reservoir with a small, deep dimple. Atoms accumulate in the laser-cooled reservoir and densely populate the dimple, in which a BEC forms in steady state. After time-of-flight expansion, the BEC shows as an elliptical feature in the centre of an absorption image. The scale bar denotes 1 mm. **b**, By off-resonantly addressing the  $^3\text{P}_1$ - $^3\text{S}_1$  transition using a ‘transparency’ laser beam, we produce a strong, spatially varying light shift on the  $^3\text{P}_1$  electronic state, rendering atoms locally transparent to laser-cooling photons addressing the  $^1\text{S}_0$ - $^3\text{P}_1$  transition. This enables condensation in the protected dimple region. **c**, Schematic of the potential landscape from both reservoir and dimple trap, and of the dominant mechanisms leading to BEC atom gain and loss.

## Experiment

To continuously refill the reservoir, a stream of atoms from an 850-K oven flows through a series of spatially separated laser-cooling stages. The initial stages load a steady-state magneto-optical trap (MOT) operated on the 7.5-kHz  $^1\text{S}_0 - ^3\text{P}_1$  transition<sup>7</sup>, shown in Fig. 1a. An atomic beam of  $\mu\text{K}$  atoms is then outcoupled and guided<sup>39</sup> 37 mm to the reservoir. This long-distance transfer prevents heating of the atoms in the reservoir by laser-cooling light used in earlier cooling stages.

To slow the roughly  $10\text{-cm}\cdot\text{s}^{-1}$  atomic beam and load it into the reservoir while minimizing resonant light, we implement a Zeeman slower on the  $|m_J=0\rangle - |m_J=1\rangle$  transition. This slower uses a single, counterpropagating laser beam together with the  $0.23\text{-G}\cdot\text{cm}^{-1}$  MOT magnetic field gradient along the guide (see [Methods](#)). The  $11.5\text{-}\mu\text{K}$ -deep reservoir is produced by a horizontal 1,070-nm laser beam focused to an elliptical spot with waists  $w_y = 14.5\ \mu\text{m}$  vertically and  $w_x = 110\ \mu\text{m}$  horizontally. A  $6^\circ$  horizontal angle between the guide and the reservoir allows the decelerated atoms to be nudged into the reservoir after reaching the intersection. The atomic beam and the reservoir are radially cooled by two pairs of beams addressing the magnetically insensitive  $|m_J=0\rangle - |m_J=1\rangle$  transition.

This arrangement of traps and cooling beams leads to the loading of the reservoir with a flux  $\Phi_R = 1.4(2) \times 10^6$  atoms  $\text{s}^{-1}$  (see [Supplementary Information](#)), a radial temperature of  $T_{Rr} = 0.85(7)\ \mu\text{K}$  and an axial temperature of  $T_{Rz} = 3.0(5)\ \mu\text{K}$ . The corresponding phase-space flux is  $\kappa = (\frac{\partial \rho}{\partial t})_T = \frac{\Phi}{(2\pi)^3 k_B T^3} \propto \frac{\hbar^3 \omega_x \omega_y \omega_z}{(2\pi)^3 k_B T^3} = 5.0(2) \times 10^{-2}$  (ref. <sup>45</sup>), in which  $\hbar$  is the reduced Planck constant,  $k_B$  the Boltzmann constant and  $\omega_R/2\pi$  are the reservoir trap frequencies.

To reduce heating and loss, we use a ‘transparency’ laser beam<sup>43</sup> that renders atoms in the dimple trap transparent to near-resonant cooling light.

This beam is overlapped with the dimple and its frequency is set at 33 GHz blue detuned from the  ${}^3\text{P}_1$ - ${}^3\text{S}_1$  transition, so as to locally apply a differential light shift on the  ${}^1\text{S}_0$ - ${}^3\text{P}_1$  transition; see Fig. 1b and [Methods](#). All transitions to the  ${}^3\text{P}_1$  manifold are thereby shifted by more than 500 times the  ${}^1\text{S}_0$ - ${}^3\text{P}_1$  linewidth, whereas atoms in the  ${}^1\text{S}_0$  ground state experience a light shift of only 20 kHz. Without the transparency beam, the lifetime of a pure BEC in the dimple is shorter than 40 ms, whereas with the transparency beam, it exceeds 1.5 s (see [Methods](#)).

For a BEC to form in the dimple, the ultracold gas must exceed a critical phase-space density of order one. The dimple is produced by a vertically propagating 1,070-nm beam with a 27-μm waist focused at the centre of the reservoir. In the steady state, the  $6.9(4) \times 10^5$  atoms in the dimple are maintained at a low temperature ( $T_D = 1.08(3)$  μK) by thermalization through collisions with the  $7.3(1.8) \times 10^5$  laser-cooled atoms in the reservoir<sup>43</sup>. The dimple provides a local density boost thanks to its increased depth (7 μK) and small volume compared with the reservoir<sup>46,47,48</sup>. This leads to a sufficient phase-space density for condensation.

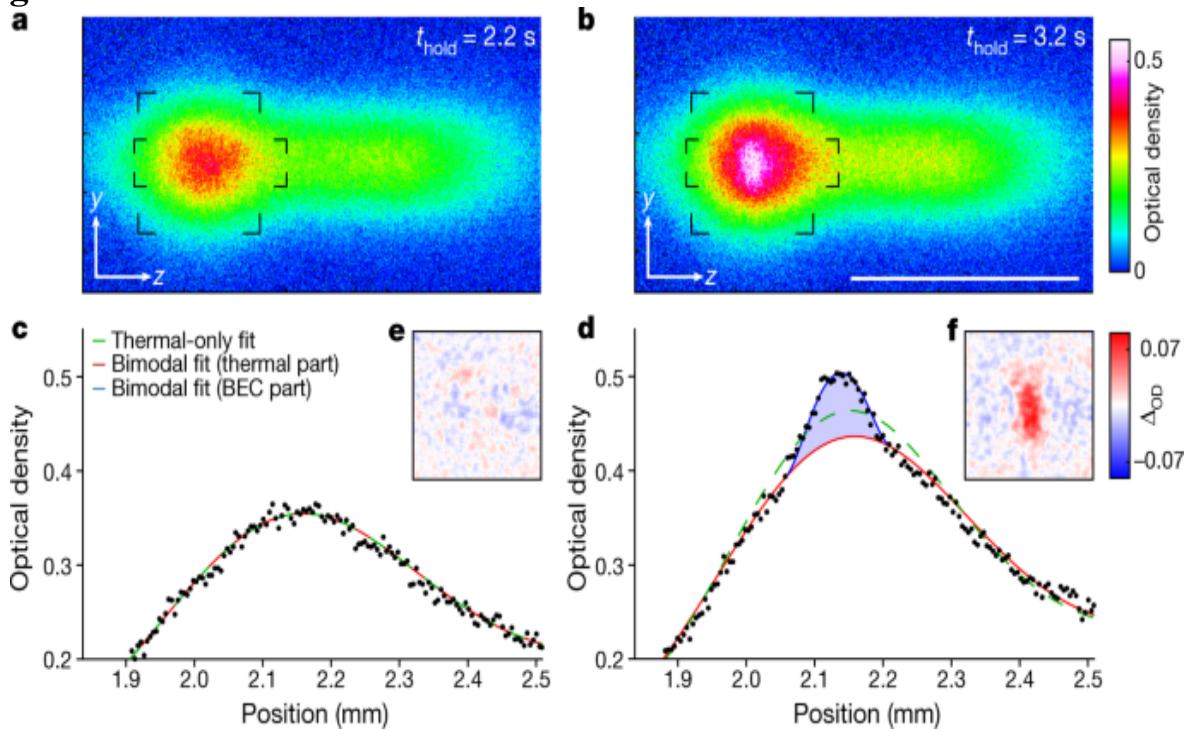
In a typical instance of our experiment, we suddenly switch all laser beams on and let atoms accumulate in the reservoir and dimple for a time  $t_{\text{hold}}$ . The phase-space density in the dimple increases and—eventually—a BEC forms. The BEC grows thanks to preferential Bose-stimulated scattering of non-condensed atoms into the macroscopically populated BEC mode. This produces continuous-matter-wave amplification, the gain mechanism for the CW BEC<sup>31</sup>. The BEC grows until losses eventually balance gain and steady state is reached.

## Analysis of the CW BEC

We now demonstrate the existence of a BEC and later show that it persists indefinitely. To tackle the first point, we analyse atomic cloud density images for  $t_{\text{hold}} = 2.2$  s and 3.2 s, immediately before and after the formation of a BEC, as shown in Fig. 2a,b. These  $x$ -integrated absorption images are taken after switching off all laser beams and letting the cloud expand for 18

ms. Both images show broad distributions of thermal atoms that are horizontally extended, reflecting the spatial distribution of the gas before expansion. Notably, the image for the longer  $t_{\text{hold}}$  shows a further small elliptical feature at the location of highest optical density, which is consistent with the presence of a BEC. The appearance of a BEC is clearly shown in Fig. 2c,d, showing  $y$ -integrated density distributions. For short  $t_{\text{hold}}$ , only a broad, thermal distribution exists. However, for long  $t_{\text{hold}}$ , a bimodal distribution appears, the hallmark of a BEC.

**Fig. 2: Detection of the CW BEC.**

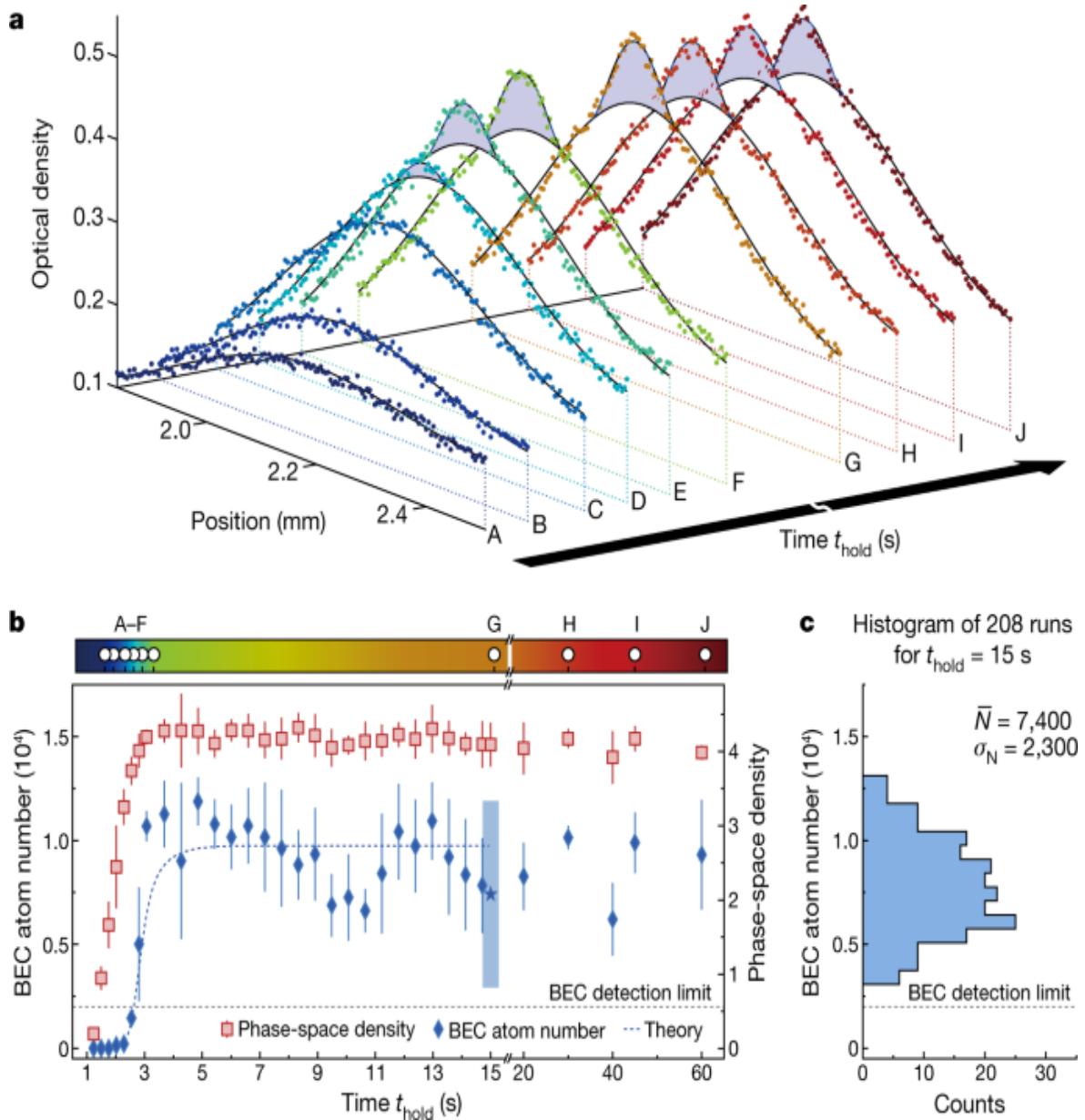


**a, b,** Absorption images of the atomic cloud before and after condensation. The atoms are imaged after an 18-ms time-of-flight expansion. The scale bar denotes 1 mm. **c, d,** Optical density within the elongated rectangles marked by corners in **a** and **b**, averaged along  $y$ . Fitted profiles using a thermal-only distribution (green dashed line) or a bimodal distribution, consisting of a thermal (red line) and a Thomas–Fermi (blue line) component. The thermal-only fit fails to represent the condensed atoms in **d** (blue shaded area). **e, f,** Corner-marked square regions of absorption images **a** and **b** minus thermal parts of the bimodal fits, showing the CW BEC.

We further validate the existence of the BEC by fitting theoretical distributions to the absorption images in Fig. 2a, b. As shown in Fig. 2c, d, excellent agreement is found by combining a thermal distribution with a Thomas–Fermi distribution describing the BEC. At short hold times, we find that a thermal fit alone is sufficient to describe the data, whereas at longer times, the extra Thomas–Fermi component is required, indicating the presence of a BEC. To clearly visualize the BEC, we remove the thermal fit component from the data; see Fig. 2e, f. The pronounced anisotropic shape of the BEC in Fig. 2f is consistent with the expansion of a BEC from the anisotropic dimple, whose trap frequency along the  $y$  axis is approximately double that along  $z$  (see [Supplementary Information](#) and Extended Data Fig. 6).

Once established, the BEC can be maintained in steady state indefinitely with gain balancing losses. As shown in Fig. 3, we study the formation transient and stability of the condensate by recording and analysing images for different  $t_{\text{hold}}$ . Figure 3a shows representative density profiles during the initial 5-s formation transient (A–F) and then in the presence of a stable BEC (G–J). Likewise, Fig. 3b shows the evolution and then stability of the BEC atom number and the peak phase-space density in the dimple,  $\langle \rho \rangle = N_D \left( \frac{\hbar^3}{\omega_x \omega_y \omega_z} \right)^{1/3} (k_B T)^{1/3}$ , in which  $N_D$  is the thermal atom number in the dimple and  $\omega_{D_i}/2\pi$  are the dimple trap frequencies. The steady-state BEC is observed over durations much longer than both the lifetime of a pure BEC (1.5–3 s) and the background-gas-limited lifetime (7 s) (see [Methods](#)).

**Fig. 3: Formation and stability of the CW BEC.**



**a**, Profiles as in Fig. 2c,d for various hold times (marked in b) first during the formation of the BEC (A: 1.5 s, B: 1.8 s, C: 2.2 s, D: 2.5 s, E: 2.8 s, F: 3.2 s) and then during the steady state of the CW BEC (G: 15 s, H: 30 s, I: 45 s, J: 60 s). **b**, Evolution of the BEC atom number and the dimple atom phase-space density  $\rho_D$  depending on hold time  $t_{\text{hold}}$  after suddenly switching on all laser beams. The blue dashed line shows the result of the BEC evolution fitted to the data before 15 s using the rate-equation model (see [Supplementary Information](#)). The error bars show the standard deviation from binning about four measurements for each time. **c**, Histogram

of the BEC atom number from 208 images for  $t_{\text{hold}} = 15$  s, long after the establishment of steady state (blue star in **b**). No points fall below our BEC detection limit of 2,000 atoms. The 95% confidence interval ( $4\sigma_N$ ) calculated from this dataset is given in **b** at 15 s (blue rectangle).

Although we do not continuously monitor the CW BEC, its atom number fluctuations can be estimated from many independent observations. To study these fluctuations, we collected about 200 measurements for  $t_{\text{hold}} = 15$  s, which is markedly longer than both the lifetimes in the system and the formation transient; see Fig. 3c. The average BEC atom number is  $\langle \bar{N} \rangle = 7,400(2,300)$ , with none of the points falling below our BEC detection threshold of 2,000 atoms (see [Methods](#)).

Modelling the formation, growth and stabilization of the BEC provides valuable insights into this new driven-dissipative system. It also provides the gain and loss from the BEC, which are important for practical applications such as producing a CW atom laser<sup>32</sup> and improving matter-wave coherence. We explain the BEC dynamics by fitting a phenomenological rate-equation model to measured temperature and atom numbers. Our analysis covers the condensate formation and perturbations such as disrupting the reservoir loading (see [Supplementary Information](#)). From this model, we estimate a steady-state gain of  $2.4(5) \times 10^5$  atoms  $s^{-1}$  into the BEC, with representative fits shown in Fig. 3b and Extended Data Fig. 7. A substantial fraction of this gain could conceivably be translated into an outcoupled flux forming a CW atom laser. We also find that losses in the BEC at steady state are dominated by three-body recombinations with thermal atoms, owing to the gas density exceeding  $5 \times 10^{14}$  atoms  $cm^{-3}$ . The presence of high, steady influx and loss makes our BEC a driven-dissipative system. We confirm this by showing that it is impossible to model the atoms in the trap as a closed system in thermal equilibrium (see [Supplementary Information](#)). Open driven-dissipative systems such as this one are thought to show rich non-equilibrium, many-body physics waiting to be explored, such as purity oscillations<sup>49</sup>, behaviours described by new critical exponents<sup>28</sup> and unusual quantum phases, especially in lower dimensions<sup>50</sup>.

## Discussion and conclusion

In summary, we have demonstrated continuous Bose–Einstein condensation. The resulting CW BEC can be sustained indefinitely using constant gain provided by a combination of Bose-stimulated scattering and atom refilling with high phase-space flux. Our work opens the door to continuous matter-wave devices. Moving forwards, many improvements are possible. In the near term, the purity of our BEC can be increased by enhancing the phase-space flux loading the dimple. A straightforward option to achieve this is to render the reservoir laser cooling uniform by using a magic-wavelength reservoir trap. Further options include lowering the reservoir temperature by Raman cooling<sup>44</sup> or by adding a continuously operating evaporation stage<sup>36</sup>. A CW BEC allows overcoming limits imposed on matter-wave coherence by the finite lifetime and atom number of a single condensate<sup>51</sup>. In practice, exceeding this limit will require extreme field stability, including external fields such as dipole trap laser fields and the condensate mean field. For example, a coherence time exceeding 1 s requires an atom number stability on the order of 0.1%. Techniques such as feedback could be used to overcome such sources of noise<sup>52,53</sup> and could ultimately allow coherence approaching the standard quantum limit or beyond<sup>51,54,55</sup>.

Our CW BEC is the matter-wave equivalent of a CW optical laser with fully reflective cavity mirrors. A tantalizing prospect is to add an output coupler to extract a propagating matter-wave. This could be implemented by coherently transferring atoms to an untrapped state and would bring the long-sought CW atom laser finally within reach<sup>15,45</sup>. This prospect is especially compelling because our CW BEC is made of strontium, the element used in some of today’s best clocks<sup>56</sup> and the element of choice for future cutting-edge atom interferometers<sup>20,21,22,23,24,57,58</sup>. Our work could inspire a new class of such quantum sensors.

## Methods

### Creating an ultracold $^{84}\text{Sr}$ beam

We use the experimental scheme developed in our previous work<sup>7,39</sup> to create an ultracold  $^{84}\text{Sr}$  beam propagating within a dipole trap guide. The scheme begins with strontium atoms emitted by an 850-K oven. They then

travel through a succession of laser-cooling stages arranged along several connected vacuum chambers using first the  $^1\text{S}_0 - ^1\text{P}_1$  and then the  $^1\text{S}_0 - ^3\text{P}_1$  transitions. Using the 30-MHz-wide  $^1\text{S}_0 - ^1\text{P}_1$  transition is necessary to efficiently slow and cool the fast atoms from the oven. However, this strong transition cannot be used in the last chamber in which the BEC is located, owing to the probable heating of the BEC from scattered near-resonant photons. Cooling using the narrow  $^1\text{S}_0 - ^3\text{P}_1$  transition is, however, made possible in this last chamber thanks to the addition of a transparency beam (see below).

To form a guided beam, atoms arriving in the final vacuum chamber are first captured and cooled in a narrow-line MOT. They are then outcoupled into a long, horizontal dipole guide with a 92- $\mu\text{m}$  waist. The  $^{84}\text{Sr}$  atoms propagate along the guide with a velocity  $v_G = 8.8(8) \text{ cm s}^{-1}$ , a Gaussian velocity spread  $\Delta v_G = 5.3(2) \text{ cm s}^{-1}$  and a flux  $\Phi_G = 8.6(1.0) \times 10^6 \text{ atoms s}^{-1}$ .

## Making the reservoir and dimple traps

The 11.5- $\mu\text{K}$ -deep reservoir is produced by a right circularly polarized 1,070-nm laser beam propagating in the  $z$  direction. It uses 540 mW of power focused to an elliptical spot with waists of  $w_y = 14.5 \mu\text{m}$  vertically and  $w_x = 110 \mu\text{m}$  horizontally. The guided atomic beam and the reservoir intersect with a horizontal angle of  $6^\circ$  about 1 mm from the reservoir centre and 37 mm from the MOT quadrupole centre. The reservoir beam crosses approximately 45(10)  $\mu\text{m}$  below the guide beam and descends with a vertical tilt of around  $1.2(1)^\circ$  as it separates from the guide beam. A secondary 250-mW beam of waist 175(25)  $\mu\text{m}$  runs parallel to the guide and points at the reservoir region. The fine adjustment of these beams is used to optimize the flow of atoms from the guide to the reservoir.

The dimple region has a 7  $\mu\text{K}$  deeper potential located at the centre of the reservoir. This is mainly produced by a vertically propagating 1,070-nm ‘dimple beam’, although 1  $\mu\text{K}$  is due to the vertically propagating transparency beam. The dimple beam uses 130 mW of power linearly polarized along the  $z$  axis with a 27- $\mu\text{m}$  waist in the plane of the reservoir. The dimple trap frequencies are  $(\omega_{Dx}, \omega_{Dy}, \omega_{Dz}) = 2\pi \times (330, 740, 315) \text{ Hz}$ ,

whereas the reservoir beam alone produces a trap with frequencies ( $\omega_{Rx}$ ,  $\omega_{Ry}$ ,  $\omega_{Rz} = 2\pi \times (95, 740, 15)$  Hz).

## Zeeman slower on the $^1S_0$ - $^3P_1$ transition

To load the guided atomic beam into the reservoir, it must first be slowed and pushed into the reservoir. To perform this task, we implement a Zeeman slower using the  $^1S_0$ - $^3P_1$  transition starting around 3 mm before the guide-reservoir intersection. The slower makes use of the quadrupole magnetic field of the narrow-line MOT to provide a magnetic gradient along the axis of the guide. The quadrupole field of the MOT has gradients of  $-0.55$ ,  $0.32$  and  $0.23$  G cm $^{-1}$  in the  $x$ ,  $y$  and  $z$  directions respectively. The slower is displaced by 37 mm along the  $z$  axis with respect to the quadrupole centre, resulting in a magnetic field offset of 0.85 G. The slower uses a counterpropagating 200- $\mu$ m-waist laser beam that crosses the guide at a shallow horizontal angle of  $4^\circ$ . We modulate the laser frequency to broaden its effective linewidth to 50 kHz. This makes the slowing robust to potential fluctuations in the effective detuning (see Extended Data Table 1). The light intensity corresponds to  $2.2 I_{sat}$  when not frequency-broadened, in which  $I_{sat} \approx 3 \mu$ W cm $^{-2}$  is the saturation intensity of the transition. We choose the laser detuning to match the Zeeman shift of the  $|(^3\{\rm{P}\}_1| |(^3\{\rm{J}\}_{\prime=1}, m_{\rm{J}} \{^3\{\rm{J}\}\})^{(\prime=-1)}$  state at the intersection between the guide and the reservoir. This way, atoms reach zero axial velocity at the intersection before being pushed back and into the reservoir.

## Loading the reservoir

Because the reservoir is a conservative trap, efficiently loading atoms from the guide requires a dissipative mechanism. This is provided in two ways by laser cooling on the  $^1S_0$ - $^3P_1$  transition. The first is a ‘counter Zeeman slower’ beam propagating approximately along the  $z$  axis opposing the Zeeman slower beam. This beam addresses the  $|(^3\{\rm{P}\}_1| |(^3\{\rm{J}\}_{\prime=1}, m_{\rm{J}} \{^3\{\rm{J}\}\})^{(\prime=-1)}$  state with a peak intensity of about  $8 I_{sat}$  and has a waist of 150  $\mu$ m. Making use of this

magnetic transition, we choose the light detuning such as to address the atoms near the guide–reservoir intersection and thus compensate the backwards acceleration of the Zeeman slower beam. This allows atoms to gradually diffuse towards the reservoir centre, in which collisions and the second laser-cooling mechanism will further lower their temperature.

The second cooling mechanism consists of a molasses on the radial axes ( $x$ ,  $y$ ) addressing the magnetically insensitive  $\pi$  transition. Using a magnetically insensitive transition avoids affecting cooling by the spatial inhomogeneities in the effective detuning owing to magnetic field variation across the extent of the laser-cooled cloud. Another cause of spatial inhomogeneities, which does affect the molasses cooling efficiency, is the differential light shift induced by the reservoir trap. This shift is around +55 kHz, many times larger than the linewidth of the transition. The optimal molasses cooling frequency is found to be 42 kHz higher than the unperturbed transition. This partially accommodates for the differential light shifts and preferentially cools atoms located near the bottom of the reservoir. To reach the lowest temperature and enable condensation in the dimple, we also apply a very low total light intensity of  $0.4 I_{\text{sat}}$ . With this choice of detuning and intensity, some of the incoming atoms reach the reservoir centre, in which they are radially cooled to  $T_{Rr} = 0.85(7) \mu\text{K}$ . Other atoms might be heated out of the 9- $\mu\text{K}$ -evaporation-threshold trap by the blue-detuned light in the outer trap region.

## Minimizing heating and loss in the reservoir

The atoms in the reservoir have a lifetime of 7 s, limited by collisions with the background gas of the vacuum chamber. However, these losses can be overwhelmed by optical effects such as photoassociation or heating by photon scattering. It is therefore critical to minimize the exposure of the reservoir to unnecessary light, and we address this point by implementing four techniques.

First, the 37-mm offset between the MOT and reservoir centres allows us to avoid any direct illumination from the  $x$ ,  $y$  MOT beams on the reservoir; see Extended Data Fig. 1. On the  $z$  axis, the influence of the MOT beams is greatly reduced by using a ‘dark cylinder’, as described in ref. [39](#).

Second, we optimize the cooling spectrum and intensity of each laser-cooling beam entering the last vacuum chamber. By separately measuring their influence on the reservoir atom number, we optimize on a compromise between the lifetime of atoms and the loading flux. The results are illustrated in Extended Data Fig. 1 and Extended Data Table 1.

Third, we maximize the  $\pi$  polarization component of the molasses beams that illuminate both the guided beam and the reservoir, thus minimizing the effects of unwanted transitions. Unavoidably, beams along the  $y$  axis possess admixtures of  $\sigma^-$  and  $\sigma^+$  owing to the orientation of the local magnetic field.

Finally, we purify the spectrum of the light used to address the  ${}^1S_0 - {}^3P_1$  cooling transition. Our cooling light is produced by several injection-locked diode lasers beginning from a single external-cavity diode laser (ECDL). We reduce the linewidth of this ECDL to 2 kHz by locking it onto a cavity with a finesse of approximately 15,000, whose spectrum has a full width at half maximum of about 100 kHz. By using the light transmitted through this cavity to injection lock a second diode laser, we can filter out the amplified spontaneous emission of the ECDL and servo bumps. This filtering is critical to increase the lifetime of the atoms inside the dimple by reducing resonant-photon scattering.

Without the dimple and transparency beams, individual laser-cooling beams reduce the lifetime of atoms in the reservoir to no shorter than about 1.5 s. With the dimple, transparency and all laser-cooling beams on, atoms in the reservoir have a 1/e lifetime of 420(100) ms, as determined from the fits shown in Extended Data Fig. 2.

## Transparency beam

To minimize the destructive effects of resonant light on the BEC and atoms within the dimple, we render this region locally transparent to light on the  ${}^1S_0 - {}^3P_1$  cooling transition. By coupling light to the  ${}^3P_1 - {}^3S_1$  transition, we induce a light shift on the  ${}^3P_1$  state, as illustrated in Extended Data Fig. 3a, b. Owing to the extreme sensitivity of the BEC to photon scattering, all sublevels of the  ${}^3P_1$  state must be shifted markedly. This requires using at least

two of the three transition types ( $\sigma^\pm, \pi$ ) in this  $J = 1 - J' = 1$  structure. However, when polarizations at the same frequency are combined, quantum interference between sub-levels always produces a dark state in the dressed  ${}^3P_1$  manifold. In this case, the energy of this dark state can only be shifted between  $\pm\Delta_{\text{Zeeman}}$ , in which  $\Delta_{\text{Zeeman}}$  is the Zeeman shift of the  $(\{\}^3\{\rm P\}_1\{m\}_{\{(\rm J)\}}^{\prime})=1$  state. This corresponds to  $\Delta_{\text{Zeeman}} = 1.78$  MHz at the dimple location, giving a light shift that is insufficient to protect the BEC. Thus it is necessary to use different frequencies for the different polarization components of the transparency beam, as illustrated in Extended Data Fig. 3c.

The transparency beam is implemented by a single beam propagating vertically and focused on the dimple location with a 23- $\mu\text{m}$  waist. This geometry aims to minimize the overlap of the transparency beam with the reservoir volume. In this way, we protect atoms at the dimple location without affecting the laser cooling taking place in the surrounding reservoir. This is necessary to maintain the high phase-space flux of the reservoir. The transparency laser light is blue detuned by 33 GHz from the 3.8-MHz-wide  ${}^3P_1 - {}^3S_1$  transition at 688 nm. This detuning is chosen to be as large as possible while still enabling sufficient light shift with the available laser power. The light contains two frequency components: 7 mW of right-hand circularly polarized light and 3 mW of left-hand circularly polarized light, separated by 1.4 GHz. The relative detuning is chosen to be large enough to avoid dark states while remaining experimentally easy to implement. It is small compared with the absolute detuning to obtain similarly good protection by each component. The relative intensity is chosen to shift all  ${}^3P_1$  states by a similar magnitude. The magnetic field at the dimple location lies in the  $(y, z)$  plane and has an angle of 60° with respect to the vertical  $y$  axis along which the transparency beam propagates. This leads to a distribution of the light intensity onto the transitions  $\{\sigma^+, \sigma^-, \pi\}$  of  $\{1, 9, 6\}$  for the left-hand and  $\{9, 1, 6\}$  for the right-hand circular polarization.

The light is produced from a single ECDL, frequency shifted by acousto-optic modulators and amplified by several injection-locked laser diodes and a tapered amplifier. Because the  ${}^1S_0 - {}^3P_1$  and  ${}^3P_1 - {}^3S_1$  lines are less than 1.5 nm apart, it is crucial to filter the light to prevent amplified spontaneous

emission from introducing resonant scattering on the  $^1S_0$ - $^3P_1$  transition. This filtering is performed by a succession of three dispersive prisms (Thorlabs PS853 N-SF11 equilateral prisms), followed by a 2.5-m (right-hand circular) or 3.9-m (left-hand circular) propagation distance before aperturing and injection into the final optical fibre.

## Characterizing the transparency beam protection

The transparency-beam-induced light shifts on the  $^1S_0$ - $^3P_1$  transition were measured spectroscopically by probing the absorption of  $^{88}\text{Sr}$  samples loaded into the dimple.  $^{88}\text{Sr}$  is used instead of  $^{84}\text{Sr}$  because the higher natural abundance improves signal without affecting the induced light shifts. Spectra are recorded for various transparency beam laser intensities at the magnetic field used for the CW BEC experiments. The results are shown in Extended Data Fig. [3c](#) for one and then two polarization components.

The observed light shifts are consistent with the calculated dressed states for the six coupled sub-levels of the  $^3P_1$  and  $^3S_1$  states. This is evaluated by solving the Schrödinger equation in the rotating frame of the light field for a transparency beam consisting of a single-frequency, right-hand circular laser beam in the presence of the measured external magnetic field. The theoretical results are given in Extended Data Fig. [3c](#) (solid lines, left side), with no adjustable parameters. We find a reasonable agreement with the observed shifts and reproduce the expected saturation of the light shift owing to the presence of a dark state. An optimized fit can be obtained with a slightly higher intensity corresponding to a waist of 21  $\mu\text{m}$  instead of 23  $\mu\text{m}$ , and a slightly modified polarization distribution. In this fitted polarization distribution, the contribution of the weakest component,  $\sigma^-$ , is enhanced by a factor of roughly 2.5. Both differences can be explained by effects from the vacuum chamber viewports and dielectric mirrors.

When the left-hand circular polarization component of the transparency beam is added, we observe in Extended Data Fig. [3c](#) (right side) that the ‘dark’ state shifts linearly away. In this manner, all sub-levels of  $^3P_1$  can be shifted by more than 4 MHz, more than 500 times the linewidth of the laser-cooling transition. For comparison, the light shift on the  $^1S_0$  ground state

from the transparency beam is 20 kHz, and at most 380 kHz by all trapping beams, about one order of magnitude smaller than the shift on  ${}^3P_1$  states from the transparency beam.

We demonstrate the protection achieved by the transparency beam in two ways. First, we measure the lifetime of a pure BEC inside the dimple in the presence of all light and magnetic fields used in the CW BEC experiments. This pure BEC is produced beforehand using time-sequential cooling stages. Once the pure BEC is produced, we apply the same conditions as used for the CW BEC, except that the light addressing the  ${}^1S_0 - {}^1P_1$  transition is off, to prevent new atoms from arriving. Without the transparency beam, the  $1/e$  lifetime of a pure BEC in the dimple cannot even reach 40 ms, whereas with the transparency beam, it exceeds 1.5 s.

Second, we show the influence of the transparency beam on the existence of a CW BEC. Beginning with the same configuration as the CW BEC but without the transparency beam, steady state is established after a few seconds, with no BEC formed. We then suddenly switch the transparency beam on and observe the evolution of the sample as shown in Extended Data Fig. 4. Although the reservoir sample seems unaffected, the dimple atom number increases by a factor of 6.4(1.8), indicating fewer losses. At the same time, the sample (partially) thermalizes and a BEC appears after about 1 s. No BEC is formed if only one transparency beam frequency component is present or only one-third of the nominal transparency beam power is applied. This demonstrates the critical importance of the transparency beam.

## Characterizing the BEC and thermal cloud

To characterize the CW BEC and surrounding thermal cloud, we switch all traps and beams off and perform absorption imaging. Fitting the distributions of the expanding clouds allows us to estimate atom numbers and temperatures throughout the system, as well as the number of condensed atoms, all from a single image.

We begin with absorption images typically recorded after an 18-ms time-of-flight expansion. The observed 2D density distribution can be fitted by an ensemble of four thermal components plus an extra Thomas–Fermi

distribution when a BEC is present. Three independent 2D Gaussian functions represent atoms originating from the dimple, the reservoir and the crossing between the guide and the reservoir. Atoms originating from the guide are represented along the guide's axis by a sigmoid that tapers off owing to the effect of the Zeeman slower and in the radial direction by a Gaussian profile. Examples are shown in Extended Data Fig. 5.

We found this fit function with 18 free parameters to be the simplest and most meaningful one capable of representing the data. By combining knowledge of their distinct locations and/or momentum spreads, we can determine individually the populations and their characteristics. We find that the uncertainty in the fitted parameters is mostly unimportant compared with shot-to-shot variations in the data. An exception is distinguishing the population in the reservoir from that in the guide–reservoir crossing region, in which there is some ambiguity, resulting in higher uncertainties. In both the main text and [Methods](#), the error bars indicate the standard deviation  $\sigma$  calculated from several images. Although it is possible to estimate the temperatures in the  $y$  axis from a single fitted image, the initial cloud sizes in the  $z$  direction are large compared with the ballistic expansion. Thus we use a set of measurements with varying times of flight to estimate  $z$ -axis temperatures.

When a BEC is present, it is necessary to add a Thomas–Fermi profile to the previously discussed fit function. The only other free parameter used in the fit is the number of atoms in the BEC. We assume that the BEC position is the same as that of the non-condensed atoms in the dimple and we calculate the radii of the BEC from the BEC atom number, the  $s$ -wave scattering length, the trap frequencies in the dimple and the expansion time<sup>59</sup>. These frequencies are calculated from the knowledge of the waists of each relevant beam and of the powers used. The waists are either directly measured or extracted from observations of dipole oscillation frequencies of a pure BEC in the trap for several beam powers.

Adding an extra fitting parameter can lead to overfitting. To rigorously determine whether including this Thomas–Fermi distribution provides a significantly better fit of the data, we use a statistical  $F$ -test. This allows us to determine a BEC atom number threshold above which the fit is statistically better than that without the Thomas–Fermi distribution. For this

$F$ -test, we isolate a region of interest (ROI) in the image containing both thermal and BEC atoms. We then calculate the value  $\langle F = \frac{\{(\{\rm{RSS}_i\})_1 - (\{\rm{RSS}_i\})_2\}^2}{\{(\{\rm{RSS}_i\})_1\} / (\{(\{\rm{RSS}_i\})_2\} / (n - p_2))} \rangle$  in which  $\rm{RSS}_i$  is the residual sum of squares over the ROI for model  $i$  with  $p_i$  parameters and  $n$  is the number of pixels of the ROI. The fit including the Thomas–Fermi distribution is significantly better than that without only if  $F$  is higher than the critical value of an  $F$ -distribution with  $(p_2 - p_1, n - p_2)$  degrees of freedom, with a desired confidence probability. By applying this test to the data of Fig. 3, we find that the BEC model fits better, with a confidence greater than 99.5%, when the BEC atom number exceeds 2,000. This sets our detection limit, above which we are confident a BEC exists. Notably, this limit is lower than the BEC atom number, corresponding to a  $-2\sigma_N$  shot-to-shot fluctuation. This shows that, at all times after steady state is reached, a BEC exists.

## Data availability

Raw data and analysis materials used in this research can be found at <https://doi.org/10.21942/uva.16610143.v1>.

## References

1. Bloch, I., Dalibard, J. & Nascimbène, S. Quantum simulations with ultracold quantum gases. *Nat. Phys.* **8**, 267–276 (2012).
2. Cronin, A. D., Schmiedmayer, J. & Pritchard, D. E. Optics and interferometry with atoms and molecules. *Rev. Mod. Phys.* **81**, 1051–1129 (2009).
3. Bongs, K. et al. Taking atom interferometric quantum sensors from the laboratory to real-world applications. *Nat. Rev. Phys.* **1**, 731–739 (2019).
4. Lachmann, M. D. et al. Ultracold atom interferometry in space. *Nat. Commun.* **12**, 1317 (2021).

5. Aguilera, D. N. et al. STE-QUEST—test of the universality of free fall using cold atom interferometry. *Class. Quantum Gravity* **31**, 115010 (2014).
6. Hartwig, J. et al. Testing the universality of free fall with rubidium and ytterbium in a very large baseline atom interferometer. *New J. Phys.* **17**, 035011 (2015).
7. Bennetts, S., Chen, C.-C., Pasquiou, B. & Schreck, F. Steady-state magneto-optical trap with 100-fold improved phase-space density. *Phys. Rev. Lett.* **119**, 223202 (2017).
8. Volchkov, V. V., Rührig, J., Pfau, T. & Griesmaier, A. Sisyphus cooling in a continuously loaded trap. *New J. Phys.* **15**, 093012 (2013).
9. Essen, L. & Parry, J. V. L. An atomic standard of frequency and time interval: a caesium resonator. *Nature* **176**, 280–282 (1955).
10. Keith, D. W., Ekstrom, C. R., Turchette, Q. A. & Pritchard, D. E. An interferometer for atoms. *Phys. Rev. Lett.* **66**, 2693–2696 (1991).
11. Kwolek, J. M., Fancher, C. T., Bashkansky, M. & Black, A. T. Three-dimensional cooling of an atom-beam source for high-contrast atom interferometry. *Phys. Rev. Appl.* **13**, 044057 (2020).
12. Savoie, D. et al. Interleaved atom interferometry for high-sensitivity inertial measurements. *Sci. Adv.* **4**, 7948 (2018).
13. Müntinga, H. et al. Interferometry with Bose-Einstein condensates in microgravity. *Phys. Rev. Lett.* **110**, 093602 (2013).
14. Ketterle, W. Nobel lecture: When atoms behave as waves: Bose-Einstein condensation and the atom laser. *Rev. Mod. Phys.* **74**, 1131–1151 (2002).
15. Robins, N. P., Altin, P. A., Debs, J. E. & Close, J. D. Atom lasers: production, properties and prospects for precision inertial measurement. *Phys. Rep.* **529**, 265–296 (2013).

16. Spreeuw, R. J. C., Pfau, T., Janicke, U. & Wilkens, M. Laser-like scheme for atomic-matter waves. *EPL* **32**, 469–474 (1995).
17. Holland, M., Burnett, K., Gardiner, C., Cirac, J. I. & Zoller, P. Theory of an atom laser. *Phys. Rev. A* **54**, 1757–1760 (1996).
18. Safronova, M. S. et al. Search for new physics with atoms and molecules. *Rev. Mod. Phys.* **90**, 025008 (2018).
19. Safronova, M. S. The search for variation of fundamental constants with clocks. *Ann. Phys. (Berl.)* **531**, 1800364 (2019).
20. Yu, N. & Tinto, M. Gravitational wave detection with single-laser atom interferometers. *Gen. Relativ. Gravit.* **43**, 1943–1952 (2010).
21. Graham, P. W., Hogan, J. M., Kasevich, M. A. & Rajendran, S. New method for gravitational wave detection with atomic sensors. *Phys. Rev. Lett.* **110**, 171102 (2013).
22. Abe, M. et al. Matter-wave atomic gradiometer interferometric sensor (MAGIS-100). *Quantum Sci. Technol.* **6**, 044003 (2021).
23. Canuel, B. et al. Exploring gravity with the MIGA large scale atom interferometer. *Sci. Rep.* **8**, 14064 (2018).
24. Badurina, L. et al. AION: an atom interferometer observatory and network. *J. Cosmol. Astropart. Phys.* **2020**, 011 (2020).
25. Lisdat, C. et al. A clock network for geodesy and fundamental science. *Nat. Commun.* **7**, 12443 (2016).
26. Grotti, J. et al. Geodesy and metrology with a transportable optical clock. *Nat. Phys.* **14**, 437–441 (2018).
27. Takamoto, M. et al. Test of general relativity by a pair of transportable optical lattice clocks. *Nat. Photonics* **14**, 411–415 (2020).
28. Sieberer, L. M., Huber, S. D., Altman, E. & Diehl, S. Dynamical critical phenomena in driven-dissipative systems. *Phys. Rev. Lett.* **110**,

195301 (2013).

29. Bhongale, S. & Holland, M. Loading a continuous-wave atom laser by optical pumping techniques. *Phys. Rev. A* **62**, 043604 (2000).
30. Chikkatur, A. P. et al. A continuous source of Bose-Einstein condensed atoms. *Science* **296**, 2193–2195 (2002).
31. Miesner, H.-J. et al. Bosonic stimulation in the formation of a Bose-Einstein condensate. *Science* **279**, 1005–1007 (1998).
32. Robins, N. P., Figl, C., Jeppesen, M., Dennis, G. R. & Close, J. D. A pumped atom laser. *Nat. Phys.* **4**, 731–736 (2008).
33. Deng, L. et al. Four-wave mixing with matter waves. *Nature* **398**, 218–220 (1999).
34. Inouye, S. et al. Phase-coherent amplification of atomic matter waves. *Nature* **402**, 641–644 (1999).
35. Schneble, D. et al. The onset of matter-wave amplification in a superradiant Bose-Einstein condensate. *Science* **300**, 475–478 (2003).
36. Lahaye, T. et al. Evaporative cooling of a guided rubidium atomic beam. *Phys. Rev. A* **72**, 033411 (2005).
37. Olson, S. E., Mhaskar, R. R. & Raithel, G. Continuous propagation and energy filtering of a cold atomic beam in a long high-gradient magnetic atom guide. *Phys. Rev. A* **73**, 033622 (2006).
38. Kindt, L. *Shock Wave Loading of a Magnetic Guide*. PhD thesis, Utrecht Univ. <https://dspace.library.uu.nl/handle/1874/211584> (2011).
39. Chen, C.-C., Bennetts, S., Escudero, R. G., Pasquiou, B. & Schreck, F. Continuous guided strontium beam with high phase-space density. *Phys. Rev. Appl.* **12**, 044014 (2019).
40. Castin, Y., Cirac, J. I. & Lewenstein, M. Reabsorption of light by trapped atoms. *Phys. Rev. Lett.* **80**, 5305–5308 (1998).

41. Williams, J., Walser, R., Wieman, C., Cooper, J. & Holland, M. Achieving steady-state Bose-Einstein condensation. *Phys. Rev. A* **57**, 2030–2036 (1998).
42. Cirac, J. I., Lewenstein, M. & Zoller, P. Collective laser cooling of trapped atoms. *EPL* **35**, 647–651 (1996).
43. Stellmer, S., Pasquiou, B., Grimm, R. & Schreck, F. Laser cooling to quantum degeneracy. *Phys. Rev. Lett.* **110**, 263003 (2013).
44. Urvoy, A., Vendeiro, Z., Ramette, J., Adiyatullin, A. & Vuletić, V. Direct laser cooling to Bose-Einstein condensation in a dipole trap. *Phys. Rev. Lett.* **122**, 203202 (2019).
45. Dennis, G. R., Davis, M. J. & Hope, J. J. Quantum kinetic theory model of a continuous atom laser. *Phys. Rev. A* **86**, 013640 (2012).
46. Pinkse, P. W. H. et al. Adiabatically changing the phase-space density of a trapped Bose gas. *Phys. Rev. Lett.* **78**, 990–993 (1997).
47. Stamper-Kurn, D. M. et al. Reversible formation of a Bose-Einstein condensate. *Phys. Rev. Lett.* **81**, 2194–2197 (1998).
48. Dutta, S. & Mueller, E. J. Kinetics of Bose-Einstein condensation in a dimple potential. *Phys. Rev. A* **91**, 013601 (2015).
49. Haine, S. A., Hope, J. J., Robins, N. P. & Savage, C. M. Stability of continuously pumped atom lasers. *Phys. Rev. Lett.* **88**, 170403 (2002).
50. He, L., Sieberer, L. M. & Diehl, S. Space-time vortex driven crossover and vortex turbulence phase transition in one-dimensional driven open condensates. *Phys. Rev. Lett.* **118**, 085301 (2017).
51. Baker, T. J., Saadatmand, S. N., Berry, D. W. & Wiseman, H. M. The Heisenberg limit for laser coherence. *Nat. Phys.* **17**, 179–183 (2020).
52. Kristensen, M. A. et al. Observation of atom number fluctuations in a Bose-Einstein condensate. *Phys. Rev. Lett.* **122**, 163601 (2019).

53. Wiseman, H. M. & Thomsen, L. K. Reducing the linewidth of an atom laser by feedback. *Phys. Rev. Lett.* **86**, 1143–1147 (2001).
54. Wiseman, H. M. Light amplification without stimulated emission: beyond the standard quantum limit to the laser linewidth. *Phys. Rev. A* **60**, 4083–4093 (1999).
55. Liu, C. et al. Proposal for a continuous wave laser with linewidth well below the standard quantum limit. *Nat. Commun.* **12**, 5620 (2021).
56. Boulder Atomic Clock Optical Network (BACON) Collaboration. Frequency ratio measurements at 18-digit accuracy using an optical clock network. *Nature* **591**, 564–569 (2021).
57. Tino, G. M. et al. SAGE: a proposal for a space atomic gravity explorer. *Eur. Phys. J. D* **73**, 228 (2019).
58. El-Neaj, Y. A. et al. AEDGE: atomic experiment for dark matter and gravity exploration in space. *EPJ Quantum Technol.* **7**, 6 (2020).
59. Castin, Y. & Dum, R. Bose-Einstein condensates in time dependent traps. *Phys. Rev. Lett.* **77**, 5315–5319 (1996).

## Acknowledgements

We thank F. Famá, S. Pyatchenkova, J. Samland, S. Zhou and the workshop of FNWI, especially J. Kalwij, T. van Klingeren and S. Koot, for technical assistance. We thank V. Barbé, C. Coulais, S. Diehl, K. van Druten, D. Guéry-Odelin, W. von Klitzing, B. van Linden van den Heuvell, R. Spreeuw, J. Tan, P. Thekkepat and J. Walraven for comments on the manuscript. We thank the NWO for funding through Vici grant no. 680-47-619 and NWA Startimpuls 2 grant NWA.QUANTUMNANO.2019.002. We thank the European Research Council (ERC) for funding under project no. 615117 QuantStro. This project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement no. 820404 (iqClock project). This work was financially supported by the Dutch Ministry of Economic Affairs and Climate Policy (EZK), as part of the Quantum Delta NL programme. B.P. thanks the NWO for funding

through Veni grant no. 680-47-438 and C.-C.C. thanks Taiwan's Ministry of Education for MOE Technologies Incubation Scholarship no. 60010200068.

## Author information

### Authors and Affiliations

1. Van der Waals-Zeeman Institute, Institute of Physics, University of Amsterdam, Amsterdam, the Netherlands

Chun-Chia Chen, Rodrigo González Escudero, Benjamin Pasquiou, Shayne Bennetts & Florian Schreck

2. Institute for Theoretical Physics, Institute of Physics, University of Amsterdam, Amsterdam, the Netherlands

Jiří Minář

3. QuSoft, Amsterdam, the Netherlands

Jiří Minář, Benjamin Pasquiou, Shayne Bennetts & Florian Schreck

### Contributions

C.-C.C. and S.B. built the apparatus. C.-C.C., R.G.E. and S.B. performed the investigation and data collection. C.-C.C., B.P. and S.B. analysed the data. J.M. developed the theoretical model. B.P., S.B. and F.S. supervised the project. F.S. acquired funding. All authors contributed to the manuscript.

### Corresponding author

Correspondence to [Florian Schreck](#).

### Ethics declarations

### Competing interests

The authors declare no competing interests.

## Peer review

### Peer review information

*Nature* thanks the anonymous reviewers for their contribution to the peer review of this work. [Peer reviewer reports](#) are available.

## Additional information

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

## Extended data figures and tables

### [Extended Data Fig. 1 Spectra of narrow-linewidth cooling lasers and their spatial extent.](#)

The right side represents the spectra of cooling lasers addressing the  $^1S_0 - ^3P_1$  transition (vertical red bars) with respect to the (relative) energy of the states in the  $^3P_1$  manifold, shown on the left side. The energies of these \(\{m\}\_{\{\rm J\}}^{\{\prime\}}\) states are given depending on the location along the  $z$  axis, and the horizontal black dashed lines represent their respective Zeeman shifts when atoms are located inside the reservoir. The horizontal colour bars at the bottom left show the location and spatial extent of each laser beam; see also Extended Data Table 1 for detailed beam parameters.

### [Extended Data Fig. 2 Loading of the reservoir and dimple at constant flux.](#)

We achieve a constant flux  $\Phi_G$  in the guide by switching the experiment on for 10 s without the Zeeman slower beam, until reaching a steady flow. We

then switch this beam on at time  $t = 0$ . We show the BEC atom number and the phase-space density  $\rho_D$  in the dimple (top). The blue dotted line indicates our BEC detection limit in terms of condensed atom number. We show the dimple, reservoir and ‘reservoir + crossing’ atom number (middle) and the temperature  $T_D$  in the dimple and the temperature  $T_{Ry}$  in the reservoir along the vertical axis (bottom). The dashed lines are the results from fits with exponential growth or decay, giving access to the (constant) fluxes, one-body loss rate parameters and thermalization times (see [Supplementary Information](#)). The error bars represent one standard deviation  $\sigma$  from binning, on average, six data points.

### **Extended Data Fig. 3 Light shift from the transparency beam.**

**a**, Level scheme showing laser-cooling and transparency transitions. **b**, Schematic of the potential energy landscape of reservoir and dimple for the  $^1S_0$  and  $^3P_1$  states. Atoms are rendered insensitive to the laser-cooling light by a single vertical ‘transparency’ laser beam (green arrow) containing two frequency components, one for each circular polarization. **c**, Transition energies to the three  $\{|m\rangle\}_{\{J\}}^{\{J'\}=0,\pm 1}$  Zeeman sub-levels of the  $^3P_1$  manifold, referenced to the transition at zero electric and magnetic field (black dashed line). The energy shifts are shown for a single right-hand circular (RHC) polarization (left) and with the addition of the left-hand circular (LHC) component (right). We show the solutions (solid lines) of the Schrödinger equation for the  $^3P_1$  manifold coupled by a light field with single frequency component and RHC polarization. In this case, at high laser intensities, the energy of the state originating from  $|m\rangle_{\{J\}}^{\{J'\}=-1}$  saturates, corresponding to the presence of a dark state. The vertical purple dashed lines show the operating intensities of the LHC and RHC light fields used in the CW BEC experiment, and the purple diamond is extrapolated from the data. The error bars indicate estimates of the ranges in which the light-shifted spectral lines lie.

### **Extended Data Fig. 4 Influence of the transparency beam.**

We let the experiment reach a steady state with the transparency beam off. At  $t = 0$ , we switch the beam on and observe the system's evolution. We show the BEC atom number and the phase-space density  $\rho_D$  in the dimple (top). The blue dotted line indicates our BEC detection limit in terms of condensed atom number. We show the dimple and reservoir atom number (middle) and the temperature  $T_D$  in the dimple and the temperature  $T_{Ry}$  in the reservoir along the vertical axis (bottom). Both atom number and temperature in the reservoir remain constant while the dimple loads further atoms, indicating lower losses thanks to the protecting effect of the transparency beam.

### Extended Data Fig. 5 Fitting a CW BEC.

We show absorption pictures and their respective fits for two hold times, before (left,  $t_{\text{hold}} = 2.2$  s) and after (right,  $t_{\text{hold}} = 3.2$  s) the formation of the BEC. In the top row, we show absorption pictures **a**, **f** taken after an 18-ms time-of-flight expansion. In the middle row, we show results of fits **b**, **g** to these pictures and the fit residuals **c**, **h**. The fits use a 2D density distribution fit function accounting only for a thermal cloud. By contrast, the bottom row shows both fits **d**, **i** and residuals **e**, **j** with a 2D density distribution fit function including a Thomas–Fermi distribution describing a BEC, in addition to the thermal distribution. In the presence of a BEC, the residual of the thermal-only fit **h** clearly shows a discrepancy at the BEC location, whereas the residual **j** demonstrates that the fit accounts for the BEC.

### Extended Data Fig. 6 BEC anisotropy after time of flight.

Absorption images after short (0.1 ms) (**a**, **e**) and long (18 ms) (**c**, **g**) free-expansion time of flight (TOF), for short ( $t_{\text{hold}} = 2.2$  s, **a**, **c**) and long ( $t_{\text{hold}} = 3.0$  s, **e**, **g**) trap loading times. Pictures **a** and **e** were imaged with a detuning of  $1.33\Gamma$  to avoid saturation. The regions of interest (corner-marked squares) centred around the density maximums are analysed in panels **b**, **d**, **f** and **h**, which show the transpose anisotropy of the density distribution. This representation produces a cloverleaf pattern when the atomic cloud is anisotropic (see [Supplementary Information](#)). For short  $t_{\text{hold}}$  (left), the cloverleaf pattern, which appears because of how the trap geometry initially

shapes the thermal cloud, keeps a constant sign and diminishes during the expansion of a thermalized gas sample. For long  $t_{\text{hold}}$  (right), we observe at long expansion time  $\mathbf{h}$  another smaller cloverleaf pattern of opposite sign, which is indicative of the presence of a BEC.

### **Extended Data Fig. 7 Modelling the onset and disappearance of the BEC.**

Evolution of the gas in the dimple. Left, BEC onset when loading reservoir and dimple at constant flux  $\Phi_G$  in the guide, started by Zeeman slower switch-on at  $t = -0.7$  s (same data as in Extended Data Fig. 2 but choosing a different origin for the time axis). Right, BEC disappearance after setting  $\Phi_G$  to zero by switching the Zeeman slower off. The first two rows show the temperature and atom number of the thermal atoms in the dimple. The last two rows show the BEC atom number and the estimated phase-space density  $\rho_D$  in the dimple. The solid blue line is the result of the fit of the BEC atom number with the model of Equation (10) of the [Supplementary Information](#). The grey dashed lines are the results of fits with exponential growth/decay functions, which are used as input for Equation (10). The error bars represent one standard deviation  $\sigma$  from binning about four data points. For the BEC atom number, owing to the small number of data points, the error bars can be underestimated compared with the more reliable characterization at steady state of  $\sigma_N = 2,300$  provided in the main text. The phase-space density is estimated from measurements of the atom number and temperature in the dimple in the same way as for Fig. 3.

### **Extended Data Table 1 Properties of laser beams addressing the narrow-linewidth $^1S_0$ - $^3P_1$ transition**

## **Supplementary information**

### **Supplementary Information**

This file contains Supplementary Methods to measure the reservoir loading rate, to detect the BEC anisotropy after time of flight and to model the open

dynamics of the BEC. It also contains a Supplementary Discussion of the model and its findings, as well as of the optical dipole trap loading and BEC creation dynamics.

## **Peer Review File**

## **Rights and permissions**

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

## **Reprints and Permissions**

## **About this article**

### **Cite this article**

Chen, CC., González Escudero, R., Minář, J. *et al.* Continuous Bose–Einstein condensation. *Nature* **606**, 683–687 (2022).  
<https://doi.org/10.1038/s41586-022-04731-z>

- Received: 15 September 2021
- Accepted: 06 April 2022
- Published: 08 June 2022

- Issue Date: 23 June 2022
- DOI: <https://doi.org/10.1038/s41586-022-04731-z>

## Share this article

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

[Get shareable link](#)

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

[Copy to clipboard](#)

Provided by the Springer Nature SharedIt content-sharing initiative

## [\*\*A step closer to atom lasers that stay on\*\*](#)

Nature Research Briefing 08 Jun 2022

---

This article was downloaded by **calibre** from <https://www.nature.com/articles/s41586-022-04731-z>

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

- Article
- Open Access
- [Published: 22 June 2022](#)

# Many-body theory of positron binding to polyatomic molecules

- [Jaroslav Hofierka](#),
- [Brian Cunningham](#),
- [Charlie M. Rawlins](#),
- [Charles H. Patterson](#) &
- [Dermot G. Green](#)

*Nature* volume **606**, pages 688–693 (2022)

- 398 Accesses
- 30 Altmetric
- [Metrics details](#)

## Abstract

Positron binding to molecules is key to extremely enhanced positron annihilation and positron-based molecular spectroscopy<sup>1</sup>. Although positron binding energies have been measured for about 90 polyatomic molecules<sup>1,2,3,4,5,6</sup>, an accurate ab initio theoretical description of positron–molecule binding has remained elusive. Of the molecules studied experimentally, ab initio calculations exist for only six; these calculations agree with experiments on polar molecules to at best 25 per cent accuracy and fail to predict binding in nonpolar molecules. The theoretical challenge stems from the need to accurately describe the strong many-body correlations including polarization of the electron cloud, screening of the

electron–positron Coulomb interaction and the unique process of virtual-positronium formation (in which a molecular electron temporarily tunnels to the positron)<sup>1</sup>. Here we develop a many-body theory of positron–molecule interactions that achieves excellent agreement with experiment (to within 1 per cent in cases) and predicts binding in formamide and nucleobases. Our framework quantitatively captures the role of many-body correlations and shows their crucial effect on enhancing binding in polar molecules, enabling binding in nonpolar molecules, and increasing annihilation rates by 2 to 3 orders of magnitude. Our many-body approach can be extended to positron scattering and annihilation  $\gamma$ -ray spectra in molecules and condensed matter, to provide the fundamental insight and predictive capability required to improve materials science diagnostics<sup>7,8</sup>, develop antimatter-based technologies (including positron traps, beams and positron emission tomography)<sup>8,9,10</sup>, and understand positrons in the Galaxy<sup>11</sup>.

## Main

Pioneering technological developments have enabled the trapping, accumulation and delivery<sup>8,9,10</sup> of positrons for study of their fundamental interactions with atoms and molecules<sup>1,12</sup>, and the formation, exploitation and interrogation of positronium (Ps)<sup>13,14</sup> and antihydrogen<sup>15,16</sup>. The ability of positrons to annihilate with atomic electrons forming characteristic  $\gamma$ -rays makes them a unique probe over vast length scales, giving them important use in, for example, materials science for ultrasensitive diagnostics of industrially important materials<sup>7,8</sup>, medical imaging (positron emission tomography)<sup>17</sup> and astrophysics<sup>11</sup>.

Proper interpretation of the materials science techniques and the development of next-generation antimatter-based technologies rely on an accurate understanding of the fundamental interactions of positrons with atoms and molecules. Substantial progress has been made developing ab initio theoretical understanding of positron–atom interactions<sup>1,12,18,19,20,21</sup>. Yet, for molecules, clusters and condensed matter, many basic aspects of positron–matter interactions remain poorly understood, and predictive capability is lacking. A notable example is the open fundamental problem of positron binding to molecules. Observation of energy-resolved annihilation

spectra have enabled measurement of positron binding energies (ranging from a few to a few hundred meV) for more than 90 molecules<sup>2,3,4,5,6</sup>. The majority of these (approximately 60) are nonpolar or weakly polar species, such as alkanes, aromatics, partially halogenated hydrocarbons, alcohols, formates and acetates. By contrast, ab initio calculations have been performed predominantly for strongly polar molecules<sup>1</sup> (though we note recent model calculations for polar and nonpolar molecules)<sup>22,23</sup>. Only six species have been studied both experimentally and with ab initio theory, via configuration interaction (CI)<sup>24,25,26,27,28</sup> and ‘any particle molecular orbital’ (APMO)<sup>29</sup> approaches: carbon disulfide CS<sub>2</sub>, acetaldehyde C<sub>2</sub>H<sub>4</sub>O, propanal C<sub>2</sub>H<sub>5</sub>CHO, acetone (CH<sub>3</sub>)<sub>2</sub>CO, acetonitrile CH<sub>3</sub>CN, and propionitrile C<sub>2</sub>H<sub>5</sub>CN<sup>1</sup>. For these, the sophisticated CI and APMO approaches proved deficient, greatly underestimating the experimental binding energies, agreeing to at best greater than approximately 25% (for acetonitrile, theory<sup>28</sup>:  $\varepsilon_b = 136$  meV, versus experiment<sup>5</sup>:  $\varepsilon_b = 180$  meV), and failing to predict binding in nonpolar CS<sub>2</sub> (versus experiment:  $\varepsilon_b = 75$  meV)<sup>4</sup> (see below). Also, the considerably larger positron–molecule binding energies compared to electron–molecule ones (that is, negative ion states)<sup>4,6</sup> are not quantitatively understood.

### Table 1 Calculated positron–molecule binding energies

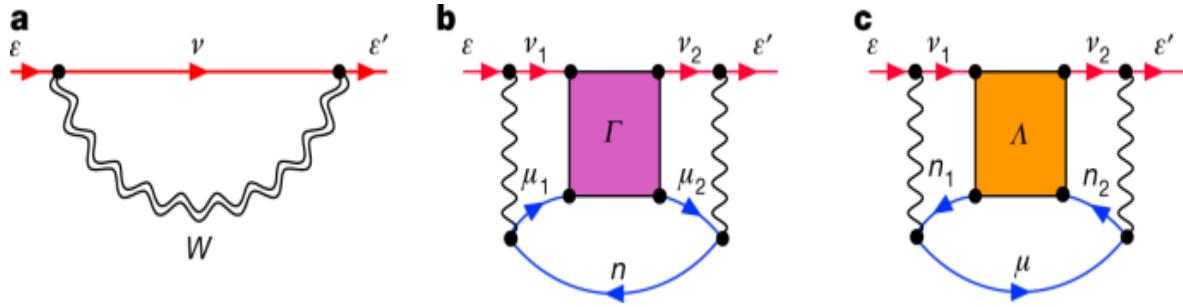
For these molecules, vibrational and geometry relaxation effects are known to provide only a few per cent correction to fixed-nuclei calculations of binding energies and wavefunction densities<sup>1,28,29,30,31,32</sup>: for example, for acetaldehyde, acetone and acetonitrile (all considered in this work) the vibrational averaging correction was approximately 1–5%<sup>28</sup>. The theoretical difficulty lies in the need to identify and accurately describe strong many-body correlations that dominate the positron–molecule system. A powerful method that can fully account for the important correlations in a natural, intuitive and systematically improvable way is many-body theory<sup>19,33,34,35,36,37,38,39</sup>.

Here we develop the many-body theory of positron interactions with polyatomic molecules. We quantify and delineate the role of the correlations including polarization of the molecular electron cloud, screening of the

positron–electron Coulomb interaction, and the unique process of virtual-Ps formation. We use the fixed-nuclei approximation and restrict to molecules with ionization energies larger than the Ps ground-state energy: for these, the Ps-formation channel is closed and the process is virtual, yet we will see that it has an important effect. After benchmarking its state-of-the-art computational implementation against explicitly correlated Gaussian (ECG) and quantum Monte Carlo (QMC) results for LiH and formaldehyde, we calculate binding energies and annihilation lifetimes for the six molecules for which both previous theory and measurements exist, finding excellent overall agreement. We additionally predict binding in formamide, CSe<sub>2</sub>, benzene and the primary nucleobases.

The positron binding energy  $\varepsilon$  and bound-state wavefunction  $\psi_\varepsilon$  is found by solving the Dyson equation<sup>33</sup>  $(H^{(0)} + \Sigma_\varepsilon)\psi_\varepsilon(\mathbf{r}) = \varepsilon\psi_\varepsilon(\mathbf{r})$ , where  $H^{(0)}$  is the Hamiltonian of the positron in the Hartree–Fock field of the ground-state molecule,  $\Sigma_E$  is a nonlocal, energy-dependent correlation potential (irreducible self energy of the positron), and  $\mathbf{r}$  is positron coordinate. It acts as an integral operator  $\Sigma_E\psi(\mathbf{r}) \equiv \int (\Sigma_E(\mathbf{r}, \mathbf{r}'))\psi(\mathbf{r}')d\mathbf{r}'$  and encapsulates the full complexity of the many-body problem. We calculate  $\Sigma$  via its expansion in residual electron–electron and electron–positron interactions, see Fig. 1. In Fig. 1a, the ‘ $GW$ ’ self energy,  $\Sigma^{GW}$ , describes the positron-induced polarization of the molecular electron cloud, and corrections to it owing to screening of the electron–positron Coulomb interaction by the molecular electrons, and electron–hole attractions (the Bethe–Salpeter equation approximation,  $GW@BSE$ ). Figure 1b represents virtual-Ps formation<sup>19,39</sup>: it is denoted by  $\Sigma^\Gamma$  and involves the summed infinite ladder series of (screened) electron–positron interactions (the ‘ $\Gamma$  block’; see Extended Data Fig. 1). The infinite ladder series is important to the positron problem because successive diagrams in this series contribute to the positron–molecule self energy with the same sign, whereas for all-electron systems the series is sign alternating and gives a small overall contribution. We also consider the smaller positron–hole ladder series contribution,  $\Sigma^A$ , shown in Fig. 1c. The construction of  $\Sigma$  and solution of the Dyson equation are detailed in [Methods](#).

**Fig. 1: Main contributions to the positron–molecule self energy.**

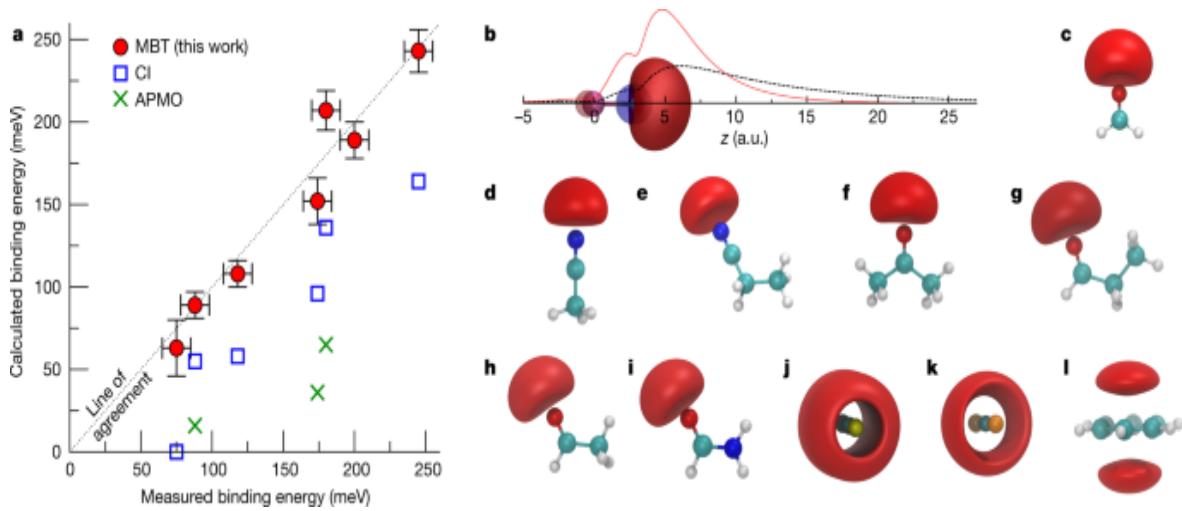


**a**, The ‘ $GW$ ’ contribution, which involves the positron Green’s function  $G_v$  and the (dynamic part of the) screened Coulomb interaction  $W$ . It describes the positron-induced polarization of the molecular electron cloud and corrections to it due to screening of the electron–positron Coulomb interaction by molecular electrons, and electron–hole attractions. **b**, The virtual-Ps contribution  $\Sigma^\Gamma$ , which includes the summed infinite ladder series (‘ $\Gamma$  block’) of screened electron–positron interactions. **c**, The positron–hole ladder series (the ‘ $\Lambda$  block’) contribution  $\Sigma^\Lambda$ . Lines directed to the right (left) represent particles (holes) propagating on the  $N$ -electron ground-state molecule: red lines labelled  $\varepsilon$  represent the external positron state; other red (blue) lines represent positron (excited electron or hole) intermediate states that are summed over; single (double) wavy lines represent bare (screened) Coulomb interactions. See Methods and Extended Data Fig. 1 for details of their calculation via the BSE.

## Positron binding energies and lifetimes

Table 1 shows our calculated binding energies at successively more sophisticated approximations to the correlation potential: Hartree–Fock,  $\Sigma^{(2)}$  (bare polarization),  $\Sigma^{GW}$  (polarization including electron screening and screened electron–hole interactions; Fig. 1a),  $\Sigma^{GW+\Gamma}$  (Fig. 1a,b), and  $\Sigma^{GW+\Gamma+\Lambda}$  (Fig. 1a–c). In the table, the first (second) number is the result using bare (dressed) Coulomb interactions in the ladders, and the third (our most sophisticated, in bold) is that using dressed interactions and energies. See also Fig. 2 for a graphical comparison of theory and experiment, and Extended Data Table 2 for more details.

**Fig. 2: Positron–molecule binding energies and bound-state Dyson wavefunction densities.**



**a**, The comparison of the present many-body calculations (red circles) with experiment (error bars for the calculations show the largest difference between the three  $\Sigma^{GW+\Gamma+A}$  calculations in Table 1). Also shown are the CI and APMO calculations (blue squares and green crosses, respectively). MBT, many-body theory. **b–l**, Positron wavefunction densities. **b**, LiH, with Li atom at origin and H at approximately 3 a.u. along the molecular axis, showing the positron wavefunction density isosurface at 70% of the maximum (red lobe), the electron HOMO wavefunction density isosurface (blue lobe is the negative region at 40% of maximum, and brown is the positive region at 10% of the maximum). Also shown is the positron wavefunction calculated along the molecular axis in the Hartree–Fock approximation (black curve) and at the  $\Sigma^{GW+\Gamma+A}$  level of many-body theory (red curve). **c–i**, The positron wavefunction density isosurfaces at 80% of maximum for formaldehyde (**c**), acetonitrile (**d**), propionitrile (**e**), acetone (**f**), propanal (**g**), acetaldehyde (**h**), and formamide (**i**). **j–l**, Nonpolar molecules with isosurfaces at 90% of maximum CS<sub>2</sub> (**j**), CSe<sub>2</sub> (**k**), and benzene (**l**). a.u., atomic units.

## [Source data](#)

## Benchmarking and general trends

We benchmark our approach against ECG ( $\varepsilon_b = 1,043$  meV)<sup>40</sup> and QMC ( $\varepsilon_b = 1,015$  meV)<sup>41</sup> calculations for LiH, and against QMC for formaldehyde ( $\varepsilon_b = 25 \pm 3$  meV)<sup>42</sup>. The LiH results demonstrate the general trends seen in

all the molecules considered. The Hartree–Fock binding energy ( $\varepsilon_b = 130$  meV) is severely deficient. Including the bare polarization attraction  $\Sigma^{(2)}$  considerably increases the binding energy (to  $\varepsilon_b = 434$  meV). The addition of short-range screening corrections reduces the polarizability and binding energy (to  $\varepsilon_b = 336$  meV, see Extended Data Table 2), but this is compensated by the inclusion of the electron–hole attractions ( $\Sigma^{GW}$ :  $\varepsilon_b = 518$  meV). This is still, however, less than half of the ECG result. The previous CI calculation<sup>24</sup> is similarly deficient. Notably, however, including the virtual-Ps formation correlation potential ( $\Sigma^{GW+\Gamma}$ ) strongly enhances the binding, more than doubling it (to  $\varepsilon_b = 1,291$  meV). Including the positron–hole ladder ( $\Sigma^{GW+\Gamma+\Lambda}$ ) slightly reduces binding (to  $\varepsilon_b = 1,106$  meV); using screened interactions in the ladders reduces it slightly further ( $\varepsilon_b = 1,038$  meV); additionally using the dressed energies in the diagram construction gives  $\varepsilon_b = 1,060$  meV, agreeing with the ECG (QMC) result to within approximately 1% (approximately 4%). For formaldehyde, the addition of virtual-Ps again drastically enhances binding (by a factor of approximately 5 over the  $GW$  result), and including the positron–hole interaction results in a binding energy  $\varepsilon_b = 28$  meV, within the error of the QMC calculation.

Comparing to our method, the ECG and QMC approaches evidently account for virtual-Ps formation to a similar accuracy, although these methods cannot be scaled to larger molecules<sup>40</sup>, and provide relatively limited insight (see below). Additionally, the correlations effect a strong localization of the positron wavefunction density at the negatively charged end of the molecule (see Fig. 2), although overall, the wavefunction is quite diffuse, asymptotically taking the form  $\psi \propto e^{-\kappa r}$  where  $\kappa = \sqrt{2/\varepsilon_b}$ . We also calculate the positron Dyson wavefunction renormalization constants  $a$  (see equation (7) in Methods and Extended Data Table 2). These represent the contribution of the positron plus molecule in the ground-state component to the bound state. Their closeness to unity suggests the picture of a positron bound to the neutral molecule (instead of a Ps atom orbiting a molecular cation)<sup>43</sup>.

## Comparison with experiment and previous theory

The best prior agreement between theory and experiment for any molecule was for acetonitrile (greater than approximately 25%). Considering the polar molecules first (Table 1 and Fig. 2), we immediately see that our full many-body theory ( $\Sigma^{GW+\Gamma+A}$ ) is superior, giving near-exact agreement (less than about 1% level) with experiment for propionitrile, propanal, acetaldehyde and formamide, and within 10% for acetonitrile and acetone (including the experimental error). (Overall we find excellent convergence in our calculation: see Methods and Extended Data Fig. 2). For all the polar molecules, the Hartree–Fock and bare ( $\Sigma^{(2)}$ ) and dressed ( $GW$ ) polarization potentials substantially underestimate binding. The effect of virtual Ps is crucial: it enhances the binding energy by a factor of approximately 2 and is essential to bring theory into agreement with experiment. We note that the previous CI and APMO ('REN-PP3', which uses a diagonal approximation and does not explicitly account for virtual-Ps formation) calculations are severely deficient.

For the nonpolar molecules, we find that binding is exclusively enabled by correlations. For  $CS_2$  a considerable binding energy of 75 meV was measured, whereas the CI calculation failed to predict binding<sup>27</sup>. We see that polarization ( $GW$ ) alone is insufficient to support binding. Notably, however, including the virtual-Ps contribution results in a prediction of large binding: our  $\Sigma^{GW+\Gamma+A}$  result of  $\varepsilon_b = 63$  meV is close to experiment. For the nonpolar molecules the positron wavefunction is delocalized around the molecule (Fig. 2j–l), making the accurate description of virtual-Ps more computationally demanding. For  $CSe_2$  and benzene, in contrast to the molecules already considered, we have not optimized the bases (accurate calculation for these molecules require computational resources currently beyond our disposal) and our values for  $\varepsilon_b$  should be considered as lower bounds. Nevertheless, the results further elucidate the essential role of virtual-Ps formation in enabling (large) binding, and the positron wavefunctions provide fundamental insight that may prove instructive to refine ab initio and model calculations (see "Predicting binding in larger molecules: nucleobases").

## Prediction for formamide

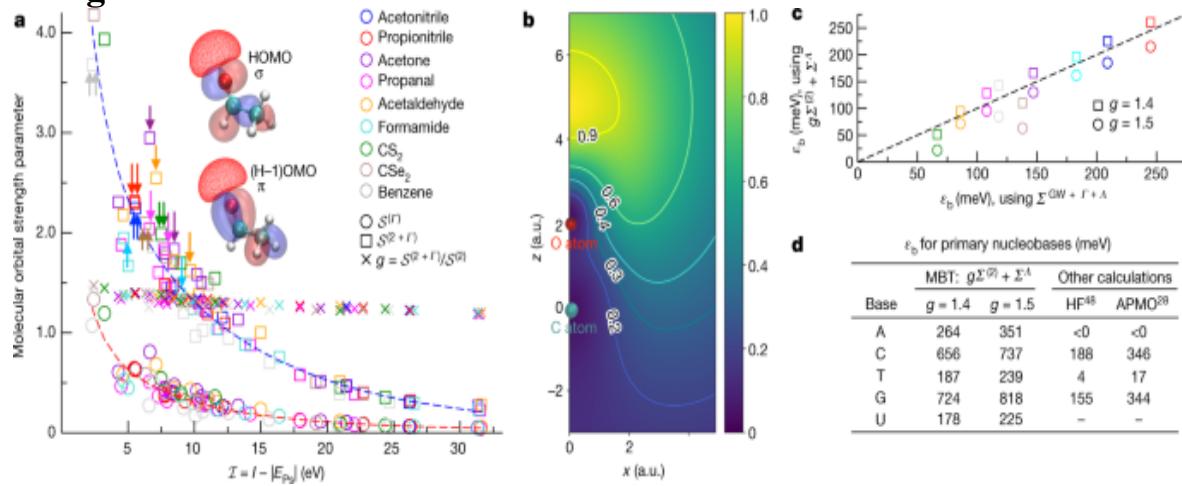
For formamide, the archetypal molecule for the investigation of protein and peptide chemistry, we are unaware of any prior calculation. We predict binding ( $\varepsilon_b \approx 189$  meV). Preliminary experiments see evidence of  $\varepsilon_b \approx 200$  meV, although a final value has yet to be determined (J. R. Danielson, S. Ghosh & C. M. Surko, unpublished material).

## Molecular orbital contributions to binding

At the static Hartree–Fock level, we find  $\varepsilon_b$  to be (monotonically and nonlinearly) related to the permanent dipole moment (expected from the dipole-potential model)<sup>44</sup>. Ultimately the correlation potential is anisotropic (see Extended Data Table 1 for calculated anisotropic polarizabilities), and depends nonlinearly on the polarizabilities and ionization energies of the individual molecular orbitals. Moreover, the binding energy depends nonlinearly on the correlation potential (for example, see Extended Data Fig. 3). The ordering of  $\varepsilon_b$  with respect to dipole moment persists to the  $\Sigma^{(2+\Gamma+\Lambda)}$  calculation, with the exception of acetaldehyde and propanal, and we note that for acetone, correlations considerably enhance  $\varepsilon_b$ . It is instructive to consider the dimensionless quantity<sup>45</sup>  $\langle (\mathbf{S}) \rangle = -\sum_{\nu} \langle \nu | \mathbf{S} | \nu \rangle / \langle \nu | \nu \rangle$  (where the sum is over excited Hartree–Fock positron basis states of energy  $\varepsilon_\nu$ , see Methods), which gives an effective measure of the strength of the correlation potential  $\Sigma$ . The magnitudes of the strength of  $\Sigma^{(2)}$ ,  $\langle (\mathbf{S}) \rangle^{(2)}$  ranges from 4–15 (see Extended Data Table 2), and follows the ordering of the isotropic polarizability, with the exception of acetone and propanal (acetone has a larger polarizability and smaller ionization energy than propanal), and benzene and CSe<sub>2</sub> (owing to benzene’s  $\pi$  bonds; see below). This suggests that (the short-range contributions to)  $\Sigma^{(2)}$  cannot be parametrized solely by the polarizability. Similarly, the magnitudes of  $\langle (\mathbf{S}) \rangle^{(\Gamma)}$  (ranging from 2–5) do not strictly follow the ordering of the ionization energies. To illuminate this, note that at the bare-polarization approximation,  $\Sigma^{(2)}$ , and polarization plus virtual-Ps formation approximation,  $\Sigma^{(2+\Gamma)} = \Sigma^{(2)} + \Sigma^{(\Gamma)}$ , we can delineate the contribution of individual molecular orbitals to positron binding. Figure 3a shows the partial  $\langle (\mathbf{S}) \rangle^{(\Gamma)}$  and  $\langle (\mathbf{S}) \rangle^{(2+\Gamma)}$  for individual occupied molecular

orbitals against their respective ionization energies, and the ratio  $\langle g \rangle_{\text{equiv}}$   $\{\{\mathcal{S}\}\}^{\{(2+\Gamma)\}}/\{\{\mathcal{S}\}\}^{\{(2)\}}$ , where  $\{\{\mathcal{S}\}\}^{\{(2)\}} = \{\{\mathcal{S}\}\}^{\{(2+\Gamma)\}} - \{\{\mathcal{S}\}\}^{\{(\Gamma)\}}$ . Both  $\langle \{\{\mathcal{S}\}\}^{\{(\Gamma)\}} \rangle$  and  $\langle \{\{\mathcal{S}\}\}^{\{(2+\Gamma)\}} \rangle$  decrease from the Ps-formation threshold to higher ionization energies: it is more difficult to perturb more tightly bound electrons. However, the decrease is not monotonic: we see that despite having larger ionization energies,  $\pi$ -type electronic molecular orbitals below the highest occupied molecular orbital (HOMO) can contribute considerably more than a  $\sigma$ -type HOMO to  $\langle \{\{\mathcal{S}\}\}^{\{(\Gamma)\}} \rangle$  and  $\langle \{\{\mathcal{S}\}\}^{\{(2+\Gamma)\}} \rangle$ —for example, in acetone, propanal and acetaldehyde, the strength of the  $\pi$ -type (H-1)OMO is larger than the  $\sigma$ -type HOMO, and in propanal, the (H-3)OMO of  $\pi$  type contributes more strongly than the (H-2)OMO, and so on. It was previously speculated<sup>3</sup> that  $\pi$  bonds were important due to the ability of the positron to more easily access electron density that is delocalized from (repulsive) nuclei. This is borne out by our calculations, and we see in Fig. 3b that considerable positron density protrudes into the region of the  $\pi$  bond. Acetonitrile and propionitrile have a doubly degenerate  $\pi$  HOMO of large strength. For acetonitrile this results in a larger strength parameter than formamide.

**Fig. 3: Molecular orbital contributions to binding, and scaling formula for large molecules.**



**a**, Molecular orbital contribution to the dimensionless strength of the virtual-Ps formation correlation potential  $\langle \{\{\mathcal{S}\}\}^{\{(\Gamma)\}} \rangle$

(circles) and including bare polarization  $\langle \{ \mathscr{S} \} \rangle^{(2+\{\Gamma\})}$  (squares) and the ratio  $\langle g \equiv \{ \mathscr{S} \} \rangle^{(2+\{\Gamma\})}/\langle \{ \mathscr{S} \} \rangle^{(2)}$  (crosses, where  $\langle \{ \mathscr{S} \} \rangle^{(2)} = \langle \{ \mathscr{S} \} \rangle^{(2+\{\Gamma\})} - \langle \{ \mathscr{S} \} \rangle^{((\{\Gamma\}))}$ ) against  $\langle I \rangle$ , where  $I$  is the molecular orbital ionization energy and  $E_{Ps} = -6.8$  eV is the ground-state energy of Ps. Arrows on  $\langle \{ \mathscr{S} \} \rangle^{(2+\{\Gamma\})}$  mark  $\pi$  orbitals with  $I < 15$  eV, and inset molecular orbital plots show HOMO ( $\sigma$ -type) and next HOMO ( $\pi$ -type) in acetaldehyde (solid red and blue: positive and negative regions of electronic molecular orbital; red wireframe: positron density at 85% of maximum). Dashed lines show the fits  $\langle \{ \mathscr{S} \} \rangle \approx a \langle e \rangle^{(-bI + c)} + d$  with  $a = 0.67$  (2.57);  $b = 0.121$  (0.092);  $c = 2.51$  (6.64);  $d = 1.38$  (1.37) for  $\langle \{ \mathscr{S} \} \rangle^{(\{\Gamma\})}$  ( $\langle \{ \mathscr{S} \} \rangle^{(2+\{\Gamma\})}$ ). **b**, The positron wavefunction density in acetaldehyde (in the plane containing the CO bond perpendicular to the CCO plane), which protrudes along the  $\pi$  bond. **c**, Comparison of binding energies for the molecules in **a** calculated using  $\Sigma = \Sigma^{GW+\Gamma+A}$  and accounting for (the computationally demanding to calculate)  $\Sigma^\Gamma$  via  $\Sigma = g\Sigma^{(2)} + \Sigma^A$ , for  $g = 1.4$  (circles) and 1.5 (squares) (see text). **d**,  $\varepsilon_b$  in nucleobases, calculated with  $\Sigma \approx g\Sigma^{(2)} + \Sigma^A$  for  $g = 1.4$  and  $g = 1.5$ . HF, Hartree–Fock; MBT, many-body theory.

## [Source data](#)

### Predicting binding in larger molecules: nucleobases

The ratio  $\langle g \equiv \{ \mathscr{S} \} \rangle^{(2+\{\Gamma\})}/\langle \{ \mathscr{S} \} \rangle^{(2)}$  depends weakly on the ionization energy, with a value of approximately 1.4–1.5 for the HOMOs ( $I \approx 10$  eV). We propose that binding energies of large molecules (for example, 15–100 atoms, for which a converged calculation of the virtual-Ps diagram (Fig. 1b) may be too computationally demanding) can be calculated by approximating  $\Sigma \approx g\Sigma^{(2)} + \Sigma^A$ . As well as accounting for virtual-Ps formation, this model potential reflects the anisotropy of the true interactions. For the molecules considered in Table 1, this works well (see Fig. 3c and Extended Data Fig. 3). Using this approximation, we calculate the positron binding energy in the five

primary nucleobases (Fig. 3d). Our results are larger than the previous APMO calculations, mirroring the results for the molecules in Table 1. Notably, we predict binding in adenine.

## Positron lifetimes

We also calculate the annihilation lifetime of the bound positron (see Methods and Extended Data Figs. 4, 5), finding that the correlations reduce it by approximately 2–3 orders of magnitude to  $\tau \approx 1$  ns. The partial annihilation rates on individual occupied electron orbitals are also calculated and are shown to depend strongly on the symmetry relative to that of the positron molecular orbital, with the HOMO not necessarily dominating, for example in acetonitrile and formamide. Such partial annihilation rates are required to properly interpret materials science experiments—for example, positron-annihilation-induced Auger electron spectroscopy<sup>7,8</sup>—and define the initial cationic wavepackets in positron-annihilation-induced charge migration<sup>46</sup>, relevant to, for example, base-selective oxidization of nucleobases<sup>47</sup>.

## Future perspectives

Many-body theory has elucidated the long-standing correlation-dominated problem of positron binding to molecules. Benchmarking against ECG and QMC calculations for LiH and formaldehyde showed the many-body theory to be similarly accurate, but its power is that it can be extended to large molecules and provides additional fundamental insight. Importantly, the effects of correlations were quantified and delineated. In particular, the key role of virtual-Ps formation in greatly enhancing binding in polar molecules and exclusively enabling binding in nonpolar molecules, the near-cancellation of screening corrections to the bare polarization, and the non-negligible role of the positron–hole interaction were all identified. The contribution of individual molecular orbitals to the (anisotropic) correlation potential was quantified, and the importance of electronic  $\pi$  orbitals (previously speculated)<sup>3</sup> was confirmed. For polar molecules the many-body theory gave binding energies in excellent (near exact, or within small error bars for most cases) agreement with the long-standing experiments. For

nonpolar molecules, binding was predicted for  $\text{CS}_2$ ,  $\text{CSe}_2$  and benzene, with larger discrepancy (within less than approximately 20%) owing to slower convergence of the virtual-Ps contribution due to the delocalized positron wavefunction. We also predicted binding in formamide and the primary nucleobases. The excellent level of agreement with experiment reaffirms that for these molecules binding is dominated by correlations, and that vibrational effects are relatively small<sup>1,28,29,30,31,32</sup>. Complementary laser-assisted photorecombination experiments<sup>48</sup> would provide direct comparison with our results, elucidating the problem in the absence of vibrations. Coupled-cluster<sup>34</sup> and ADC<sup>36</sup> calculations should also be possible and instructive.

The present calculations support resonant annihilation experiments and the related theory to which binding energies and annihilation lifetimes enter as parameters<sup>1</sup>. Beyond the fundamental insight immediately provided, the step-change in capability enables calculations of positron scattering and molecular-orbital-resolved, Doppler-broadened annihilation  $\gamma$ -ray spectra (underway), required to, for example, properly interpret positron-based ultrasensitive materials science techniques<sup>7,8</sup>, provide insight on molecular fragmentation<sup>10</sup>, and understand positron interactions in the Galaxy<sup>11</sup> and living tissue (relevant to developing next-generation PET)<sup>49,50</sup>. It also provides a foundation for the implementation of inelastic<sup>38</sup> (to include real Ps formation) and time-dependent<sup>51</sup> molecular processes, to, for example, model positron-annihilation-induced Auger-electron spectroscopy<sup>8</sup>, interatomic Coulomb decay<sup>52</sup>, charge migration<sup>46</sup> (including that relevant to medicine)<sup>47</sup> and luminescence<sup>53</sup>, and for the study of vibrational effects via coupling of the nuclear and electronic degrees of freedom<sup>54</sup>. Finally, the difficulty of the positron–molecule problem presents a rich testbed for the development of other approaches to the many-body problem, for which our results can serve as benchmarks.

## Methods

### The Dyson equation in a Gaussian basis

We calculate the positron–molecule binding energy  $\varepsilon$  and quasiparticle wavefunction  $\psi_\varepsilon$  by solving the Dyson equation. We take the zeroth-order Hamiltonian  $H^{(0)}$  to be that of the positron in the Hartree–Fock field of the frozen-target  $N$ -electron ground-state molecule. The self-energy diagrams thus begin at second order in the Coulomb interaction. Instead of computing the self energy  $\Sigma(\mathbf{r}, \mathbf{r}')$  in the coordinate basis, it is more convenient to work with its matrix elements in the Hartree–Fock basis. Specifically, we expand the electron (−) and positron (+) Hartree–Fock molecular orbitals  $\langle\langle\phi_i\rangle\rangle$  in distinct Gaussian basis sets as  $\langle\langle\phi_i\rangle\rangle = \sum_{k=1}^N \sum_{A} \langle\langle N_A | c_k^\dagger | \chi_A(k) \rangle\rangle$ , where  $A$  labels the  $\langle\langle N_A | c_k^\dagger | \chi_A(k) \rangle\rangle$  basis centres,  $k$  labels the  $\langle\langle N_A | c_k^\dagger | \chi_A(k) \rangle\rangle$  different Gaussians on centre  $A$ , each taken to be of Cartesian type with angular momentum  $L^x + L^y + L^z$ , and with  $\langle\langle\chi_A(k)|\chi_A(k')\rangle\rangle = \langle\langle \mathcal{N}_A \rangle\rangle \langle\langle A_k | k' \rangle\rangle \langle\langle x_A | x_{A'} \rangle\rangle^x \langle\langle y_A | y_{A'} \rangle\rangle^y \langle\langle z_A | z_{A'} \rangle\rangle^z \exp\{-\zeta_A | \mathbf{r} - \mathbf{r}' |\}$ , where  $\langle\langle \mathcal{N}_A \rangle\rangle$  is a normalization constant, and  $C$  are the expansion coefficients to be determined (see later in this section). Molecular geometries are determined via minimization of the total electronic Hartree–Fock energy in the Molpro<sup>56,57</sup> package, for the respective basis set (see next paragraph), ensuring an internally consistent ab initio calculation.

For both electrons and positrons, we use the diffuse-function-augmented correlation-consistent polarized aug-cc-pVXZ ( $X = T$  or  $Q$ ) Dunning basis sets centred on all atomic nuclei of the molecule, which enables accurate determination of the electronic structure including cusps<sup>58</sup> and expulsion of the positron density from the nuclei. To capture the long-range correlation effects, for the positron we also additionally include at least one large even-tempered set at the molecular centre or region of maximum positron density of the form  $Ns(N-1)p(N-2)d(N-3)f(N-4)g$  with  $N \approx 10-15$  (where it should be understood that the full degenerate set of nonzero angular momentum functions is used) and exponents  $\zeta_{A_1} = \beta^{k-1}$ ,  $k = 1, \dots, N$ , for each angular momentum type, where  $\zeta_{A_1} > 0$  and  $\beta > 1$  are parameters. The value of  $\zeta_{A_1}$  is important because the bound positron

wavefunction behaves asymptotically as  $\psi \propto e^{-\kappa r}$ , where  $\kappa = \sqrt{2/\varepsilon}$ . Thus, to ensure that the expansion describes the wavefunction well at  $r \approx 1/\kappa$ —that is, that the broadest Gaussian covers the extent of the positron wavefunction—one must have  $|\zeta_A| \lesssim \kappa^2 = 2/\varepsilon$ . In practice we performed binding energy calculations for a range of  $|\zeta_A|$  and  $\beta$  for each molecule, finding that there are broad ranges of stability. The optimal  $|\zeta_A|$  was typically found to be in the range of  $10^{-4}$ – $10^{-3}$  for *s*- and *p*-type Gaussians and  $10^{-3}$ – $10^{-2}$  for *d*- and *f*-type Gaussians, whereas *g*-type Gaussian exponents usually had  $|\zeta_A| = 10^{-1}$  (atomic units are assumed throughout unless otherwise specified). The optimal  $\beta$  ranges from 2.2 to 3.0 depending on the number of functions  $N$  in a given shell. Finally, to simultaneously describe the expulsion of the positron wavefunction from the nuclei, and accurately describe positron density maxima away from the molecule, we strategically place additional (H atom) aug-cc-pVXZ (X = T, Q) basis sets on ‘ghost’ centres close to the regions of maximum positron density. These additional Gaussians provide additional flexibility in the basis, enabling a better description of the positron wavefunction. In particular, they play an important role in enabling the accurate description of virtual-Ps formation, which occurs several atomic units away from the molecule, and which would require high angular momentum functions to resolve the electron–positron distance in Ps away if a single basis centre was used<sup>19</sup>. By placing Gaussian basis functions of angular momentum (we use  $l \leq 4$ ) on multiple ghost centres, higher angular momentum functions can be effectively generated in the interstitial regions (see appendix B of ref. <sup>59</sup> for details). In practice, for each molecule, we perform calculations with successively increasing number of ghost centres, whose locations are optimized manually until the binding energy stabilizes. The optimum locations are found iteratively: the calculation with zero ghosts generally indicates the region of maximum positron density, around which the ghosts are targeted in subsequent calculations, thus improving the density. As a general rule, we found that the optimal distance of the ghost centres from the atom closest to the maximum of positron density is about 1 Å. For some larger or nonpolar molecules, we use multiple ghost centres surrounding the molecule. To check convergence with respect to the number and location of these ghost centres, for each molecule we performed calculations including TZ or QZ

bases on a successively increasing number of ghost centres in different arrangements until the increase in binding energy fell below a few per cent. We found that including ghosts can increase binding energies by  $\sim 10\%$  in the polar molecules, and easily by  $\sim 30\%$  for the nonpolar ones—for example, for  $\text{CS}_2$  we obtained  $\varepsilon_b = 39 \text{ meV}$  at  $GW@\text{BSE}+\Gamma+A$  level with no ghosts, rising to  $\varepsilon_b = 68 \text{ meV}$  with 16 additional ghosts. The use of higher angular momenta and more ghosts could be expected to further increase the binding energies of the nonpolar molecules. We also investigated the difference of using aug-cc-pVXZ for  $X = T, Q$  in the atomic-centred and ghost bases, and higher angular momenta in the even-tempered basis. Some improvement was noted moving from  $X = T$  to  $Q$ , and also when  $g$  states were included in addition to  $f$ , to a level of a 5%–10% in polar molecules, and 10%–30% in nonpolar molecules. Overall, good convergence with respect to both the electron and positron bases was observed (see for example, Extended Data Fig. 2).

The coefficients  $C$  in the expansion of the positron wavefunction in Gaussians are found by solving the Roothaan equations  $F^\pm C^\pm = S^\pm C^\pm \varepsilon^\pm$ , where  $F^\pm$  is the Fock matrix and  $S$  is the overlap matrix. The one-body and two-body Coulomb integrals of the Fock matrix are calculated using the McMurchie–Davidson algorithm<sup>60</sup>. We eliminate linearly dependent states by excluding eigenvalues  $< 10^{-5}$  of the overlap matrices (typically  $\leq 5\%$  of the states). In practice, to minimize the basis dimensions we transform all quantities to a spherical harmonic Gaussian basis (for a given angular momentum, the number of Cartesian Gaussians is greater than or equal to the number of spherical harmonic Gaussians)<sup>61</sup>. Solution of the Roothaan equations yield bases of electron and positron Hartree–Fock molecular orbitals  $\langle (\{\phi\}_{\alpha})^{\pm}(\{bf{r}\}) \rangle$  (which include ground and other negative energy states, and discretized continuum states) with which the self-energy diagrams can be constructed (see the next section for details).

Expanding the positron Dyson wavefunction in the positron Hartree–Fock molecular orbital basis as  $\langle (\{\psi\}_{\nu}) (\{\varphi\}_{\nu}) \rangle = \sum_{D} \langle (\{\rm{r}\}_{\nu})^{\pm} (\{\varphi\}_{\nu}) \rangle$  transforms the Dyson equation to the linear matrix equation  $HD = \varepsilon D$ , where  $\langle (\{\nu\}_{1}) | H | (\{\nu\}_{2}) \rangle = \langle \varphi_{\nu_1} | \varphi_{\nu_2} \rangle$

$\{\delta\}_{\{\nu\}_1\{\nu\}_2} + \langle \nu_1 | \{\Sigma_{\varepsilon}\}_{\{\nu\}_1\{\nu\}_2} | \nu_2 \rangle$ ). Note that we calculate the full self-energy matrix including off-diagonal terms. Such a non-perturbative approach is essential for nonpolar molecules, where binding is enabled exclusively by correlations. In practice, to obtain the self-consistent solution to the Dyson equation, we calculate the self energy at a number of distinct energies  $E_i$  spanning the true binding energy  $\varepsilon_b$ , with the latter determined from the intersection of the  $\varepsilon_b(E_i)$  data with the line  $\varepsilon_b(E) = E$ .

## The positron–molecule self energy

As discussed in the main text (Fig. 1), we consider three contributions to the irreducible self energy of the positron in the field of the molecule:  $\Sigma^{GW}$  (which describes polarization, screening and electron–hole interactions);  $\Sigma^\Gamma$  (which describes the non-perturbative process of virtual-Ps formation); and  $\Sigma^A$  (which includes the infinite ladder series of positron–hole interactions). In practice, we construct the individual contributions by first solving the respective Bethe–Salpeter equations (BSE; see Extended Data Fig. 1) for the electron–hole polarization propagator  $\Pi$ , the two-particle positron–electron propagator  $\langle G \rangle_{\{\rm II\}}^{\{\rm ep\}}$  and the positron–hole two-‘particle’ propagator<sup>33</sup>  $\langle G \rangle_{\{\rm ph\}}^{\{\rm ph\}}$ . Their general form is  $L(\omega) = L^{(0)}(\omega) + L^{(0)}(\omega)KL(\omega)$ , where the  $L^{(0)}$  are non-interacting two-body propagators and  $K$  are the interaction kernels<sup>33,62,63</sup> (for example, see Extended Data Fig. 1e for the BSE for the electron–hole polarization propagator  $\Pi$ ). In the excitation space of pair product Hartree–Fock orbitals  $L = (C\omega - H)^{-1} = \xi(\omega - \Omega)^{-1}\xi^{-1}C^{-1}$ , where the pair transition amplitudes  $\xi$  are the solutions of the pseudo-Hermitian linear-response generalized eigenvalue equations<sup>63,64,65</sup>  $H\xi = C\xi\Omega$ ,  $\xi^\dagger C\xi = C$ , where

$$\begin{aligned} \text{H} &= (\begin{array}{cc} A & B \\ B^\dagger & A^\dagger \end{array}) \\ \text{x} &= (\begin{array}{cc} X & Y \\ Y^\dagger & X^\dagger \end{array}) \\ C &= (\begin{array}{cc} 1 & 0 \\ 0 & -1 \end{array}) \\ \Omega &= (\begin{array}{cc} \Omega_+ & 0 \\ 0 & \Omega_- \end{array}), \end{aligned} \quad (1)$$

for excitation energies  $\langle\{\{\Omega\}\}_{+}^{\alpha}\rangle$  and  $\langle\{\{\Omega\}\}_{-}^{\alpha}\rangle$ , which are labelled by  $\alpha = 1, \dots, \dim(A)$ . Here the  $A$  and  $B$  matrices depend on the particular two-particle propagator  $L$  under consideration and the approximation used for it (see Extended Data Table 4 for the explicit matrix elements): note that  $B = 0$  for the two-particle propagators involving the positron, because the vacuum state for the diagrammatic expansion is that of the  $N$ -electron molecule, and thus there are no positron holes and only time-forward positron propagators. To determine the amplitudes, we use the parallel diagonalization algorithm of a previous work<sup>66</sup>, which exploits a similarity transform that gives the eigenvalues of  $C^{-1}H$  as the square roots of the eigenvalues of  $(A + B)(A - B)$  (thus requiring matrices of dimension of the  $A$  block, that is, half of the full BSE matrix dimension) to obtain  $\langle X = \frac{1}{2}(L_2 U + L_1 V) \{\{\Omega\}\}_{+}^{-1/2} \rangle$  and  $\langle Y = \frac{1}{2}(L_2 U - L_1 V) \{\{\Omega\}\}_{+}^{-1/2} \rangle$ , via the Cholesky decompositions  $\langle A+B = L_1 L_1^T \rangle$  and  $\langle A-B = L_2 L_2^T \rangle$ , and the singular value decomposition  $\langle L_2 L_1^T \rangle = U \{\Omega\} V^T \rangle$ , where  $T$  indicates the transpose. The positron-molecule self-energy matrix elements can then be written as:

$$\begin{aligned}
 & \langle\langle \nu_1 | \Sigma_E^G | \nu_2 \rangle\rangle = \sum_{\alpha} \frac{w_1 \{\nu_1\} \{\nu_3\}^{\alpha} \langle P_i \alpha | w_2 \{\nu_2\} \{\nu_3\}^{\alpha} | P_i \alpha \rangle}{w_2 \{\nu_2\} \{\nu_3\}^{\alpha}} \{E - \rm{varepsilon}_n\}_{\nu_3} - \langle\langle \nu_1 | \Omega_{+}^{\alpha} | \nu_2 \rangle\rangle \eta_{\alpha}, \\
 & (2) \\
 & \langle\langle \nu_1 | \Sigma_E^{\Gamma} | \nu_2 \rangle\rangle = \sum_{\alpha} \frac{w_1 \{\nu_1\} n^{\alpha} \langle \Gamma_{\alpha} | w_2 \{\nu_2\} n^{\alpha} | \Gamma_{\alpha} \rangle}{w_2 \{\nu_2\} n^{\alpha}} \{E - \langle\langle \nu_1 | \Omega_{+}^{\alpha} | \nu_2 \rangle\rangle\} + \langle\langle \nu_1 | \rm{varepsilon}_n | \nu_2 \rangle\rangle, \\
 & (3) \\
 & \langle\langle \nu_1 | \Sigma_E^{\Lambda} | \nu_2 \rangle\rangle = \sum_{\alpha} \frac{w_1 \{\nu_1\} \mu^{\alpha} \langle \Lambda_{\alpha} | w_2 \{\nu_2\} \mu^{\alpha} | \Lambda_{\alpha} \rangle}{w_2 \{\nu_2\} \mu^{\alpha}} \{E - \langle\langle \nu_1 | \Omega_{+}^{\alpha} | \nu_2 \rangle\rangle\} - \langle\langle \nu_1 | \Sigma_E^{\Gamma} | \nu_2 \rangle\rangle - \langle\langle \nu_1 | \Sigma_E^{\Lambda} | \nu_2 \rangle\rangle,
 \end{aligned}$$

(4)

where  $v_1, v_2$  and  $v_3$  denote positron indices and  $\mu$  and  $n$  denote electron excited states and holes respectively,  $\Sigma^{(2)}$ —which results from the  $\Pi^{(0)}$  contribution to  $\Sigma^{GW}$  and is present in both  $\langle\{G\}_{\{\{\rm II\}\}}^{\{\{\rm ep\}\}}\rangle$  and  $\langle\{G\}_{\{\{\rm II\}\}}^{\{\{\rm ph\}\}}\rangle$ —is subtracted to prevent double counting, and

$$\begin{aligned} \text{\$\$} & \begin{array}{c} \{c\} \\ \{w\}_{\{\{\nu\}_{\{1\}}\{\nu\}_{\{3\}}\}}^{\{\{\Pi,\alpha\}\}} = \sum \\ \{\mu n\}(\{\nu\}_{\{1\}}\{\nu\}_{\{3\}}|\mu n)(X)_{\{\mu n\}}^{\{\Pi,\alpha\}} + \\ \{Y\}_{\{\mu n\}}^{\{\Pi,\alpha\}}, \\ \{w\}_{\{\{\nu\}_{\{1\}}n\}}^{\{\{\Gamma,\alpha\}\}} = \sum \\ \{\mu\{\nu\}_{\{3\}}\}(\{\nu\}_{\{1\}}n|\{\nu\}_{\{3\}}\mu)X_{\{\{\nu\}_{\{3\}}\mu\}}^{\{\Gamma,\alpha\}}, \\ \{w\}_{\{\{\nu\}_{\{1\}}\mu\}}^{\{\{\Lambda,\alpha\}\}} = \sum \\ \{n\{\nu\}_{\{3\}}\}(\{\nu\}_{\{1\}}\mu|\{\nu\}_{\{3\}}n)X_{\{\{\nu\}_{\{3\}}n\}}^{\{\Lambda,\alpha\}}, \end{array} \\ & \text{\$\$} \end{aligned} \quad (5)$$

where chemists' notation for Coulomb matrix elements ( $v_1 v_3 | \mu n$ ) and so on is used (see Extended Data Table 4). The total self energy is calculated as  $\Sigma = \Sigma^{GW} + \Sigma^F + \Sigma^A$ . Such addition of the individual channels is routine in atomic many-body theory calculations<sup>19,67,68</sup> and in condensed matter, for example, the fluctuation-exchange ('FLEX') approximation<sup>69,70,71</sup>. We note that the above approach is restricted to molecules with ionization energies larger than the ground-state energy of Ps (6.8 eV). For these, the Ps-formation channel is closed, and Ps formation proceeds as a virtual process (with the electron temporarily tunnelling to the positron). For molecules for which the ionization is smaller than the energy of ground-state Ps, the inelastic Ps-formation channel is open. The above approach does not account for such inelastic channels. We note, however, that there are methods proposed to include inelastic channels in a many-body formalism<sup>38</sup>. Its implementation is beyond the scope of this paper, but would be a worthwhile future endeavour.

We implement the above in the massively parallelized EXCITON+ code developed by us, adapting the EXCITON code<sup>72,73,74</sup> to include positrons. EXCITON employs density-fitting (of the electronic density) methods<sup>74,75,76,77,78,79</sup> in a Gaussian-orbital basis for calculation of the

electronic self-energy and four-centre integrals that appear in the  $A$  and  $B$  matrices of the BSE for finite<sup>73</sup> and periodic<sup>72,74</sup> systems. The EXCITON+ code developed at Queen's University Belfast adapts EXCITON to additionally solve the positron–molecule Hartree–Fock problem, construct the full (nondiagonal) positron–molecule self energy (calculating  $w^{\Pi}$ ,  $w^T$  and  $w^A$  via density fitting of the electronic density, and including screening terms in the ladders), and solve the Dyson equation and calculate the positron–electron contact density (lifetime with respect to annihilation). The use of density fitting reduces four-centre Coulomb integrals to products of three-centre Coulomb integrals and matrix elements of the Coulomb operator between atomic orbital basis functions. Thus, the memory scaling is approximately  $\mathcal{O}(\{N\}_{-}^{-2}\{M\}_{-})$ , where  $\mathcal{O}(\{N\}_{-})$  is the total number of electron basis functions, and  $\mathcal{O}(\{M\}_{-})$  is the number of electron auxiliary basis functions. The most computationally demanding part of our approach is in the calculation of the virtual-Ps self-energy contribution  $\Sigma^T$ . For this,  $\dim A = \dim X^T = N_{\nu} \times N_{\mu}$ , the product of total number of positron molecular orbitals and excited electron molecular orbitals. For the calculations considered here,  $N_{\nu}$  ranged from 400–500 and  $N_{\mu}$  from 300–400, resulting in  $\dim X^T = 120,000–200,000$ ; thus, diagonalizing the matrix of  $(\dim X^T)^2$  elements demanded between ~100 GB and 1.5 TB of random access memory (RAM). The calculations were performed on two AMD EPYC 128 CPU @ 2 GHz, 768 GB RAM nodes of the United Kingdom Tier-2 supercomputer ‘Kelvin-2’ at Queen's University Belfast. By contrast, the  $GW$  calculations involve  $\dim A = \dim X^{\Pi} \leq N_{\nu} \times N_n$ , that is, a maximum equal to the product of the number of occupied and excited electron molecular orbitals. In practice, not all occupied orbitals need to be included because the tightly bound lowest occupied molecular orbitals (LOMOs) are less susceptible to perturbation by the positron and have negligible contribution to the self energy. Thus, because  $N_n \ll N_{\mu} < N_{\nu}$ , ab initio  $GW@RPA/TDHF/BSE$  calculations (RPA, random phase approximation; TDHF, time-dependent Hartree–Fock; BSE, Bethe–Salpeter equation) are considerably less computationally expensive, and can be performed for molecules or clusters with ~100 atoms, providing at least lower bounds on the positron binding energies. Moreover, as discussed (see Fig. 3c and Extended Data Fig. 3) and demonstrated for nucleobases (Fig. 3d), the virtual-Ps formation contribution can be approximated by scaling

the  $\Sigma^{(2)}$  self energy by the strength parameter ratio  $\langle g \rangle_{\text{equiv}}$   $\{\{\mathbf{S}\}\}^{(2+\{\Gamma\})}/\{\{\mathbf{S}\}\}^{(2)}$ , namely  $\Sigma \approx g\Sigma^{(2)} + \Sigma^A$ , thus enabling computationally relatively inexpensive binding-energy calculations that account for virtual-Ps formation for molecules of  $\sim 100$  atoms. Ab initio calculations for larger molecules including the virtual-Ps self energy will be feasible with additional computational resources, as would calculations using different truncated product spaces of excited electron and positron molecular orbitals and extrapolating to the basis set limit.

## Improving the accuracy of calculations

As mentioned in the previous section, the computationally intensive calculations presented here were performed using relatively modest computational resources. Access to national supercomputing facilities would enable more complete basis sets and further exploration of the effect of ghost basis centres. Numerical accuracy can also be systematically improved in a number of ways. Exploiting the molecular point group symmetry via symmetry-adapted bases and using integral screening techniques would improve the efficiency of the calculations, enabling more complete basis sets to be used. This would ultimately improve the description of the correlations (particularly in generating higher angular momenta for improved description of the virtual-Ps formation process). The calculation of the positron-molecule self energy can be improved by implementing a self-consistent diagram approach in which the positron-molecule self energy is built from  $GW$  calculated electron and positron Dyson orbitals instead of Hartree-Fock ones<sup>[33,80](#)</sup>, and/or by coupling the three self-energy channels  $\Sigma^{GW}$ ,  $\Sigma^\Gamma$  and  $\Sigma^A$  by approximating the three-particle propagators via the Faddeev<sup>[81](#)</sup>, parquet<sup>[69](#)</sup> or ADC(3)<sup>[36](#)</sup> methods (expected to be computationally feasible for small molecules using national supercomputing facilities). Moreover, the diagrammatic series should be amenable to a diagrammatic Monte Carlo<sup>[82,83](#)</sup> prescription, a powerful stochastic simulation method that enables the effective summation of many more (classes of) diagrams than considered here.

## Positron annihilation rate in the bound state

The solution of the Dyson equation also yields the positron bound-state wavefunction  $\psi_\varepsilon$ . Using it, the  $2\gamma$  annihilation rate in the bound state \((\{\Gamma\}=\{\rm{pi}\}\{r\}\_0^2 c \{\delta\}\_{\{\rm{ep}\}})\) ( $\Gamma[\text{ns}^{-1}] = 50.47 \delta_{\text{ep}} [\text{a.u.}]$ )—the inverse of which is the lifetime of the positron–molecule complex with respect to annihilation—can be calculated. Here  $r_0$  is the classical electron radius,  $c$  is the speed of light and  $\delta_{\text{ep}}$  is the electron–positron contact density,

$$\begin{aligned} & \sum_{n=1}^{N_e} \int |\psi(r)|^2 \gamma_n dr \\ & (6) \end{aligned}$$

Here the sum is over all  $N_e$  occupied electron molecular orbitals with wavefunctions  $\varphi_n$ , and  $\gamma_n$  are molecular-orbital-dependent enhancement factors that account for the short-range electron–positron attraction<sup>20,84</sup>. Recent many-body calculations for atoms by one of us determined them to follow a physically motivated scaling with the ionization energy<sup>20,84</sup> \((\{\gamma\}\_n = 1 + \sqrt{1.31/\{\varepsilon\}\_n} + \{(0.834/\{\varepsilon\}\_n)\}^{2.15})\) (where quantities are in a.u.), which we assume to hold here. The positron Dyson wavefunction is a quasiparticle wavefunction that is the overlap of the wavefunction of the  $N$ -electron ground state molecule with the fully correlated wavefunction of the positron plus  $N$ -electron molecule system<sup>33</sup>. It is normalized as

$$\begin{aligned} & \int |\psi(r)|^2 dr = \{(1 - \\ & \partial \varepsilon / \partial r)^{-1}\}^a \quad a < 1, \\ & (7) \end{aligned}$$

which estimates the contribution of the ‘positron plus molecule in the ground state’ component to the positron–molecule bound-state wavefunction, that is, the degree to which the positron–molecule bound state is a single-particle state, with smaller values of  $a$  signifying a more strongly correlated state. Extended Data Figs. 4, 5 present contact density data. Extended Data Fig. 4a shows the individual molecular orbit contribution to

the contact density as a function of the molecular orbit ionization energy. As in Fig. 3 (contribution of strength parameters from individual molecular orbits), overall the contact density increases as the ionization energy decreases: the positron overlap is greater with the more diffuse electronic HOMOs. However, molecular orbitals below the HOMO can in fact dominate, for example, acetonitrile, as shown further in Extended Data Fig. 5a–c, and Extended Data Fig. 6 for the primary nucleobases.

## Data availability

Additional relevant data are available at <https://doi.org/10.17034/04a9ffbe-e0c6-44e4-98bc-a2d30df3424c>. Source data are provided with this paper.

## Code availability

The results presented in this study were generated using the EXCITON+program, which was developed by the Queen’s University Belfast group, adapting the EXCITON code (authored by C.H.P.) to incorporate positrons and extending its many-body theory capabilities. The version of EXCITON+ used to generate the current results is available at <https://doi.org/10.17034/04a9ffbe-e0c6-44e4-98bc-a2d30df3424c>. The latest EXCITON and EXCITON+ source codes are currently available under an Open Source license from C.H.P. (EXCITON) and D.G.G. (EXCITON+) on reasonable request, but we intend to detail them in subsequent publications.

## References

1. Gribakin, G. F., Young, J. A. & Surko, C. M. Positron–molecule interactions: resonant attachment, annihilation, and bound states. *Rev. Mod. Phys.* **82**, 2557–2607 (2010).
2. Gilbert, S. J., Barnes, L. D., Sullivan, J. P. & Surko, C. M. Vibrational-resonance enhancement of positron annihilation in molecules. *Phys. Rev. Lett.* **88**, 043201 (2002).

3. Danielson, J. R., Young, J. A. & Surko, C. M. Dependence of positron–molecule binding energies on molecular properties. *J. Phys. B* **42**, 235203 (2009).
4. Danielson, J. R., Gosselin, J. J. & Surko, C. M. Dipole enhancement of positron binding to molecules. *Phys. Rev. Lett.* **104**, 233201 (2010).
5. Danielson, J. R., Jones, A. C. L., Gosselin, J. J., Natisin, M. R. & Surko, C. M. Interplay between permanent dipole moments and polarizability in positron–molecule binding. *Phys. Rev. A* **85**, 022709 (2012).
6. Danielson, J. R., Jones, A. C. L., Natisin, M. R. & Surko, C. M. Comparisons of positron and electron binding to molecules. *Phys. Rev. Lett.* **109**, 113201 (2012).
7. Tuomisto, F. & Makkonen, I. Defect identification in semiconductors with positron annihilation: experiment and theory. *Rev. Mod. Phys.* **85**, 1583–1631 (2013).
8. Hugenschmidt, C. Positrons in surface physics. *Surf. Sci. Rep.* **71**, 547–594 (2016).
9. Danielson, J. R., Dubin, D. H. E., Greaves, R. G. & Surko, C. M. Plasma and trap-based techniques for science with positrons. *Rev. Mod. Phys.* **87**, 247–306 (2015).
10. Fajans, J. & Surko, C. M. Plasma and trap-based techniques for science with antimatter. *Phys. Plasmas* **27**, 030601 (2020).
11. Prantzos, N. et al. The 511 kev emission from positron annihilation in the galaxy. *Rev. Mod. Phys.* **83**, 1001–1056 (2011).
12. Surko, C. M., Gribakin, G. F. & Buckman, S. J. Low-energy positron interactions with atoms and molecules. *J. Phys. B* **38**, R57–R126 (2005).
13. Brawley, S. J. et al. Electron-like scattering of positronium. *Science* **330**, 789–789 (2010).

14. Cassidy, D. B. Experimental progress in positronium laser physics. *Eur. J. Phys. D* **72**, 53 (2018).
15. Baker, C. J. et al. Laser cooling of antihydrogen atoms. *Nature* **592**, 35–42 (2021).
16. Andresen, G., Ashkezari, M., Baquero-Ruiz, M. et al. Trapped antihydrogen. *Nature* **468**, 673–676 (2010).
17. Wahl, R. L. & Beanlands, R. S. B. *Principles and Practice of PET and PET/CT* (Lippincott Williams and Wilkins, 2008).
18. Harabati, C., Dzuba, V. A. & Flambaum, V. V. Identification of atoms that can bind positrons. *Phys. Rev. A* **89**, 022517 (2014).
19. Green, D. G., Ludlow, J. A. & Gribakin, G. F. Positron scattering and annihilation on noble-gas atoms. *Phys. Rev. A* **90**, 032712 (2014).
20. Green, D. G. & Gribakin, G. F.  $\gamma$ -Ray spectra and enhancement factors for positron annihilation with core electrons. *Phys. Rev. Lett.* **114**, 093201 (2015).
21. Kadyrov, A. S. & Bray, I. Recent progress in the description of positron scattering from atoms using the convergent close-coupling theory. *J. Phys. B* **49**, 222002 (2016).
22. Swann, A. R. & Gribakin, G. F. Positron binding and annihilation in alkane molecules. *Phys. Rev. Lett.* **123**, 113402 (2019).
23. Sugiura, Y. et al. Positron–electron correlation-polarization potential model for positron binding in polyatomic molecules. *J. Comp. Chem.* **41**, 1576–1585 (2020).
24. Strasburger, K. Quantum chemical study on complexes of the LiH molecule with  $e^+$ , Ps and  $Ps^-$  including correlation energy. *Chem. Phys. Lett.* **253**, 49–52 (1996).
25. Tachikawa, M., Kita, Y. & Buenker, R. J. Bound states of the positron with nitrile species with a configuration interaction multi-component

molecular orbital approach. *Phys. Chem. Chem. Phys.* **13**, 2701–2705 (2011).

26. Tachikawa, M., Kita, Y. & Buenker, R. J. Bound states of positron with simple carbonyl and aldehyde species with configuration interaction multi-component molecular orbital and local vibrational approaches. *New J. Phys.* **14**, 035004 (2012).
27. Koyanagi, K., Takeda, Y., Oyamada, T., Kita, Y. & Tachikawa, M. Positron attachment to nonpolar or small dipole CXY (X, Y = O, S, and Se) molecules: vibrational enhancement of positron affinities with configuration interaction level of multi-component molecular orbital approach. *Phys. Chem. Chem. Phys.* **15**, 16208–16213 (2013).
28. Tachikawa, M. Positron attachment to acetonitrile, acetaldehyde, and acetone molecules: vibrational enhancement of positron affinities with configuration interaction level of multi-component molecular orbital approach. *J. Phys. Conf. Ser.* **488**, 012053 (2014).
29. Romero, J., Charry, J. A., Flores-Moreno, R., Varella, M. Td. N. & Reyes, A. Calculation of positron binding energies using the generalized any particle propagator theory. *J. Chem. Phys.* **141**, 114103 (2014).
30. Gianturco, F. A. et al. Positron binding to alkali-metal hydrides: the role of molecular vibrations. *Phys. Rev. A* **73**, 022705 (2006).
31. Buenker, R. J. & Liebermann, H.-P. Configuration interaction calculations of positron binding to molecular oxides and hydrides and its effect on spectroscopic constants. *Nucl. Instrum. Meth. B* **266**, 483–490 (2008).
32. Buenker, R. J., Liebermann, H.-P., Pichl, L., Tachikawa, M. & Kimura, M. Role of the electric dipole moment in positron binding to the ground and excited states of the BeO molecule. *J. Chem. Phys.* **126**, 104305 (2007).

33. Dickhoff, W. H. & Van Neck, D. *Many-body Theory Exposed! Propagator Description of Quantum Mechanics in Many-Body Systems* 2nd edn (World Scientific, 2008).
34. Bartlett, R. J. & Musiał, M. Coupled-cluster theory in quantum chemistry. *Rev. Mod. Phys.* **79**, 291–352 (2007).
35. Amusia, M. Y., Cherepkov, N. A., Chernysheva, L. V. & Shapiro, S. G. Elastic scattering of slow positrons by helium. *J. Phys. B* **9**, L531–L534 (1976).
36. Schirmer, J., Cederbaum, L. S. & Walter, O. New approach to the one-particle Green's function for finite Fermi systems. *Phys. Rev. A* **28**, 1237–1259 (1983).
37. Müller, M. & Cederbaum, L. S. Many-body theory of composite electronic-positronic systems. *Phys. Rev. A* **42**, 170–183 (1990).
38. Cederbaum, L. S. Optical potentials for elastic and inelastic scattering of non-electronic projectiles from electronic targets. *Few-Body Syst.* **21**, 211–225 (1996).
39. Dzuba, V. A., Flambaum, V. V., Gribakin, G. F. & King, W. A. Many-body calculations of positron scattering and annihilation from noble-gas atoms. *J. Phys. B* **29**, 3151–3175 (1996).
40. Bubin, S. & Adamowicz, L. Non-Born–Oppenheimer study of positronic molecular systems:  $e^+LiH$ . *J. Chem. Phys.* **120**, 6051–6055 (2004).
41. Kita, Y., Maezono, R., Tachikawa, M., Towler, M. & Needs, R. J. Ab initio quantum Monte Carlo study of the positronic hydrogen cyanide molecule. *J. Chem. Phys.* **131**, 134310 (2009).
42. Yamada, Y., Kita, Y. & Tachikawa, M. Theoretical prediction of the binding of a positron to a formaldehyde molecule using a first-principles calculation. *Phys. Rev. A* **89**, 062711 (2014).

43. Mitroy, J., Bromley, M. W. J. & Ryzhikh, G. G. Positron and positronium binding to atoms. *J. Phys. B* **35**, R81–R116 (2002).
44. Gribakin, G. F. & Swann, A. R. Effect of dipole polarizability on positron binding by strongly polar molecules. *J. Phys. B* **48**, 215101 (2015).
45. Dzuba, V. A. & Gribakin, G. F. Correlation-potential method for negative ions and electron scattering. *Phys. Rev. A* **49**, 2483–2492 (1994).
46. Kuleff, A. I. & Cederbaum, L. S. Ultrafast correlation-driven electron dynamics. *J. Phys. B* **47**, 124002 (2014).
47. Koyanagi, K., Kita, Y., Shigeta, Y. & Tachikawa, M. Binding of a positron to nucleic base molecules and their pairs. *Chem. Phys. Chem.* **14**, 3458–3462 (2013).
48. Surko, C. M., Danielson, J. R., Gribakin, G. F. & Continetti, R. E. Measuring positron–atom binding energies through laser-assisted photorecombination. *New J. Phys.* **14**, 065004 (2012).
49. Pichl, L., Tachikawa, M., Buenker, R., Kimura, M. & Rost, J.-M. The effects of positron binding and annihilation mechanisms in biomolecules on PET resolution. *IEEE Trans. Nucl. Sci.* **52**, 2810–2817 (2005).
50. Schippers, S. et al. Roadmap on photonic, electronic and atomic collision physics: II. Electron and antimatter interactions. *J. Phys. B* **52**, 171002 (2019).
51. Pavlyukh, Y., Perfetto, E. & Stefanucci, G. Photoinduced dynamics of organic molecules using nonequilibrium Green’s functions with second-Born, *GW*, *T*-matrix, and three-particle correlations. *Phys. Rev. B* **104**, 035124 (2021).
52. Sisourat, N., Miteva, T., Gorfinkel, J. D., Gokhberg, K. & Cederbaum, L. S. Interatomic Coulombic electron capture from first principles. *Phys. Rev. A* **98**, 020701 (2018).

53. Stenson, E. V., Hergenhahn, U., Stoneking, M. R. & Pedersen, T. S. Positron-induced luminescence. *Phys. Rev. Lett.* **120**, 147401 (2018).
54. Brand, J., Cederbaum, L. S. & Meyer, H.-D. Dynamical Green's function and an exact optical potential for electron–molecule scattering including nuclear dynamics. *Phys. Rev. A* **60**, 2983–2999 (1999).
55. Johnson, R. D. III (ed.) *NIST Computational Chemistry Comparison and Benchmark Database* NIST standard reference database number 101, release 21 August 2020 (NIST, accessed 1st July 2021);  
<https://doi.org/10.18434/T47C7Z>
56. Werner, H.-J. et al. The Molpro quantum chemistry package. *J. Chem. Phys.* **152**, 144107 (2020).
57. Eckert, F., Pulay, P. & Werner, H.-J. Ab initio geometry optimization for large molecules. *J. Comput. Chem.* **18**, 1473–1483 (1997).
58. Kato, T. On the eigenfunctions of many-particle systems in quantum mechanics. *Commun. Pure Appl. Math.* **10**, 151–177 (1957).
59. Swann, A. R. & Gribakin, G. F. Calculations of positron binding and annihilation in polyatomic molecules. *J. Chem. Phys.* **149**, 244305 (2018).
60. McMurchie, L. E. & Davidson, E. R. One- and two-electron integrals over Cartesian Gaussian functions. *J. Chem. Phys.* **26**, 218–231 (1978).
61. Schlegel, H. B. & Frisch, M. J. Transformation between Cartesian and pure spherical harmonic Gaussians. *Int. J. Quantum Chem.* **54**, 83–87 (1995).
62. Fetter, A. L. & Walecka, J. D. *Quantum Theory of Many-Particle Systems* (Dover, 2003).
63. Linderberg, J. & Öhrn, Y. *Propagators in Quantum Chemistry* 2nd edn (Wiley, 2004).

64. Ring, P. & Schuck, P. *The Nuclear Many-Body Problem* (Springer, 1980).
65. Dreuw, A. & Head-Gordon, M. Single-reference ab initio methods for the calculation of excited states of large molecules. *Chem. Rev.* **105**, 4009–4037 (2005).
66. Shao, M., da Jornada, F. H., Yang, C., Deslippe, J. & Louie, S. G. Structure preserving parallel algorithms for solving the Bethe–Salpeter eigenvalue problem. *Linear Algebra Appl.* **488**, 148–167 (2016).
67. Boyle, J. & Pindzola, M. *Many-Body Atomic Physics* (Cambridge Univ. Press, 1998).
68. Gribakin, G. F. & Ludlow, J. Many-body theory of positron–atom interactions. *Phys. Rev. A* **70**, 032720 (2004).
69. Martin, R. M., Reining, L. & Ceperley, D. *Interacting Electrons* (Cambridge Univ. Press, 2016).
70. Bickers, N. & Scalapino, D. Conserving approximations for strongly fluctuating electron systems. I. Formalism and calculational approach. *Ann. Phys.* **193**, 206–251 (1989).
71. Bickers, N. E., Scalapino, D. J. & White, S. R. Conserving approximations for strongly correlated electron systems: Bethe–Salpeter equation and dynamics for the two-dimensional Hubbard model. *Phys. Rev. Lett.* **62**, 961–964 (1989).
72. Patterson, C. H. Exciton: a code for excitations in materials. *Mol. Phys.* **108**, 3181–3188 (2010).
73. Patterson, C. H. Photoabsorption spectra of small Na clusters: TDHF and BSE versus CI and experiment. *Phys. Rev. Mater.* **3**, 043804 (2019).
74. Patterson, C. H. Density fitting in periodic systems: application to TDHF in diamond and oxides. *J. Chem. Phys.* **153**, 064107 (2020).

75. Whitten, J. L. Coulombic potential energy integrals and approximations. *J. Chem. Phys.* **58**, 4496–4501 (1973).
76. Dunlap, B. I., Connolly, J. W. D. & Sabin, J. R. On the applicability of LCAO-X $\alpha$  methods to molecules containing transition metal atoms: the nickel atom and nickel hydride. *Int. J. Quantum Chem.* **12**, 81–87 (1977).
77. Dunlap, B. I., Connolly, J. W. D. & Sabin, J. R. On some approximations in applications of X $\alpha$  theory. *J. Chem. Phys.* **71**, 3396–3402 (1979).
78. Baerends, E., Ellis, D. & Ros, P. Self-consistent molecular Hartree–Fock–Slater calculations I. The computational procedure. *Chem. Phys.* **2**, 41–51 (1973).
79. Vahtras, O., Almlöf, J. & Feyereisen, M. Integral approximations for LCAO-SCF calculations. *Chem. Phys. Lett.* **213**, 514–518 (1993).
80. Kotani, T., van Schilfgaarde, M. & Faleev, S. V. Quasiparticle self-consistent  $GW$  method: a basis for the independent-particle approximation. *Phys. Rev. B* **76**, 165106 (2007).
81. Degroote, M., Van Neck, D. & Barbieri, C. Faddeev random-phase approximation for molecules. *Phys. Rev. A* **83**, 042517 (2011).
82. Houcke, K. V., Kozik, E., Prokof'ev, N. & Svistunov, B. Diagrammatic Monte Carlo. *Phys. Procedia* **6**, 95–105 (2010).
83. Prokof'ev, N. & Svistunov, B. Bold diagrammatic Monte Carlo technique: when the sign problem is welcome. *Phys. Rev. Lett.* **99**, 250201 (2007).
84. Green, D. G. & Gribakin, G. F. Enhancement factors for positron annihilation on valence and core orbitals of noble-gas atoms. In *Concepts, Methods and Applications of Quantum Systems in Chemistry and Physics* (eds. Wang, Y. et al.) 243–263 (Springer, 2018).

85. Haynes, W. M. (ed.) *CRC Handbook of Chemistry and Physics* 97th edn (CRC Press, 2016).

## Acknowledgements

We thank J. Danielson and C. M. Surko (University of California, San Diego) for providing unpublished data for formamide, and A. Elena and M. Plummer (STFC Scientific Computing Laboratory, Daresbury, UK), and I. Stewart and L. Fernandez Menchero (Queen's University Belfast (QUB)) for high-performance computing support. D.G.G. additionally thanks G. Gribakin for discussions and for drawing his attention to this problem, L. S. Cederbaum for encouragement and bringing a number of references to our attention, and A. Swann, M. Grüning, S. Gregg, J. Cassidy, M. J. Gregg, S. J. Smartt and M. Borghesi for comments on the manuscript. This work was supported by the European Research Council grant 804383 ‘ANTI-ATOM’ (D.G.G.).

## Author information

### Author notes

1. These authors contributed equally: Jaroslav Hofierka, Brian Cunningham

## Authors and Affiliations

1. School of Mathematics and Physics, Queen's University Belfast, Belfast, UK

Jaroslav Hofierka, Brian Cunningham, Charlie M. Rawlins & Dermot G. Green

2. School of Physics, Trinity College Dublin, Dublin, Ireland

Charles H. Patterson

## Contributions

B.C. and D.G.G. led, supported by J.H., C.M.R. and C.H.P., the development of the positron–molecule many-body theory and its computational implementation in the newly developed EXCITON<sup>+</sup> code, which was based on the EXCITON code developed by C.H.P. for electronic structure calculations. J.H. performed the majority of calculations, supported by C.M.R. D.G.G. led the analysis, supported by J.H. D.G.G. additionally conceived and supervised the work and drafted the manuscript, which all authors contributed to editing.

## Corresponding author

Correspondence to [Dermot G. Green](#).

## Ethics declarations

## Competing interests

The authors declare no competing interests.

## Peer review

## Peer review information

*Nature* thanks Masanori Tachikawa and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

## Additional information

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

## Extended data figures and tables

[Extended Data Fig. 1 The main contributions to the positron–molecule self energy, including the two-particle propagators.](#)

**a**, The  $GW$  diagram involves the positron Green's function  $G_v$  and the dynamic part (due to the absence of an electron–positron exchange interaction) of the screened Coulomb interaction  $W_d = v\boldsymbol{\Pi}v$  (bold denotes operator form), where  $\boldsymbol{\Pi}$  is the electron–hole polarization propagator (see **b**). It satisfies the BSE (**e**) with kernel  $K = v - W_{RPA}$  (**f**), where  $W_{RPA} = v + W_{d,RPA}$  is the screened electron–hole Coulomb interaction calculated in the random phase approximation (RPA). Setting  $K = 0$  results in the bare polarization entering  $W$  only, and gives the  $\Sigma^{(2)}$  approximation, so-called as it is a second-order diagram in the electron–positron Coulomb interaction. Setting  $K = v$ , the direct part of the Coulomb interaction only, gives the random phase approximation ( $GW@RPA$ ). Setting  $K = v - v_{exch}$ —that is, including exchange which gives rise to interactions within the bubbles—yields the ‘time-dependent Hartree–Fock’ approximation ( $GW@TDHF$ ). Using screened Coulomb interactions in the exchange term is the Bethe–Salpeter approximation ( $GW@BSE$ ). **c**, The virtual-positronium contribution including the summed infinite ladder series of screened electron–positron interactions (‘ $\Gamma$  block’) shown in **g**. The contribution **d** contains the  $\Lambda$  block that is the ladder series of positron–hole interactions, which satisfies a linear integral equation of the same form as that shown in **g**.

### Extended Data Fig. 2 Convergence of positron binding energies in acetonitrile and CSe<sub>2</sub> with respect to electron and positron basis size.

Positron binding energy calculated using the  $\Sigma^{GW@BSE+\Gamma+\Lambda}$  self energy for varying number of electron (positron) Hartree–Fock molecular orbitals (whose energies are shown as blue and red crosses, respectively) included in the basis. For acetonitrile, the varying electron (positron) molecular orbital calculations included all positron (electron) molecular orbitals. For CSe<sub>2</sub>, the varying electron molecular orbital calculations included all positron molecular orbitals, whereas the varying positron molecular orbital calculation included 113 electron molecular orbitals (indicated by the lowest blue circle). The binding energy reaches convergence when the electronic orbital with energies up to ~150–200 eV are included. Similar behaviour was also observed for the other molecules considered.

[Source data](#)

**Extended Data Fig. 3 Nonlinearity of the binding energy and strength of correlation potential.**

Binding energy calculated approximating the positron self energy  $\Sigma$  as  $\Sigma \approx g\Sigma^{(2)} + \Sigma^4$  as a function of the scaling parameter  $\backslash(g\backslash\text{equiv}\{\{\mathscr{S}\}\}^{(2+\Gamma)} / \{\{\mathscr{S}\}\}^{(2)}\})$  (circles; see text for more details). Experiment (squares) is from refs. [4,5](#); for formamide preliminary measurements find a binding energy of  $\varepsilon_b \approx 200$  meV, but a final result is yet to be determined. See also Fig. [3c](#).

[Source data](#)

**Extended Data Fig. 4 Calculated electron–positron contact density.**

**a**, Contact density for individual electronic molecular orbitals as a function of their ionization energy, calculated including vertex enhancement factors at  $GW@\text{BSE}+\Gamma+A$  level (see equation [\(6\)](#)). Red dashed line: positronium ground state energy at  $|E_{\text{Ps}}| = 6.8$  eV. Grey line:  $\delta_{\text{ep}} = 0.008/(I - |E_{\text{Ps}}|)$  (for a guide). Also see Extended Data Table [3](#) and Extended Data Fig. [5](#). **b**, Contact density calculated at the Hartree–Fock (circles) and various levels of many-body theory (diamonds:  $GW@\text{BSE}$ ; squares:  $GW@\text{BSE}+\Gamma+A$ ) against the square root of the binding energy. For nonpolar systems the *s*-type bound-state contact density is expected to show linear dependence (see equation 36 of the methods in ref. [1](#)). For polar molecules there is no strict theoretical basis for this, although it is interesting that the data show an approximate linear dependency.

[Source data](#)

**Extended Data Fig. 5 Calculated electron–positron contact densities and positron lifetimes with respect to annihilation.**

**a**, Fractional contribution of individual molecular orbitals to the total electron–positron contact density (equation (6)). **b, c**, The electron–positron contact density (magenta) at the  $\Sigma^{GW+\Gamma+A}$  level for the (doubly degenerate) HOMO and (H–2)OMO in acetonitrile (blue and brown show negative and positive electron wavefunction regions, respectively), compare with Fig. 2c (positron density). **d**, Positron lifetimes with respect to annihilation.  $\tau^{(0)}$ : lifetime calculated in the Hartree–Fock independent particle approximation excluding the vertex enhancement factors and using a positron wavefunction normalized to unity;  $\tau^{GW}$  and  $\tau$ : lifetime calculated using the Dyson positron wavefunction at the  $\Sigma^{GW}$  and  $\Sigma^{GW+\Gamma+A}$  levels including vertex enhancement factors and renormalization constants (equations (6), (7)).

[Source data](#)

### Extended Data Fig. 6 Calculated electron–positron contact densities and positron lifetimes with respect to annihilation for nucleobases.

**a**, Fractional contribution of individual molecular orbitals to the total electron–positron contact density (equation (6)). **b**, Table of positron lifetimes with respect to annihilation.  $\tau^{(0)}$ : lifetime calculated in the Hartree–Fock independent particle approximation excluding the vertex enhancement factors and using a positron wavefunction normalized to unity;  $\tau^{GW}$  and  $\tau^{MBT}$ : lifetime calculated using the Dyson positron wavefunction at the  $\Sigma^{GW}$  and  $g\Sigma^{(2)} + \Sigma^{(4)}$  levels for  $g = 1.4$  and  $1.5$  including vertex enhancement factors and renormalization constants (equations (6), (7)).

[Source data](#)

### **Extended Data Table 1 Calculated polarizabilities (in Å<sup>3</sup>) and ionization energies (in eV)**

**Extended Data Table 2 Positron binding energies in the  $GW$  approximation (meV), dimensionless correlation-potential strength parameters and Dyson wavefunction renormalization constants  $a$**

**Extended Data Table 3 Positron–molecule annihilation contact densities (a.u.)**

## **Extended Data Table 4 Matrix elements of the Bethe–Salpeter linear response Hamiltonian**

### **Source data**

[\*\*Source Data Fig. 2\*\*](#)

[\*\*Source Data Fig. 3\*\*](#)

[\*\*Source Data Extended Data Fig. 2\*\*](#)

[\*\*Source Data Extended Data Fig. 3\*\*](#)

[\*\*Source Data Extended Data Fig. 4\*\*](#)

[\*\*Source Data Extended Data Fig. 5\*\*](#)

[\*\*Source Data Extended Data Fig. 6\*\*](#)

### **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

Hofierka, J., Cunningham, B., Rawlins, C.M. *et al.* Many-body theory of positron binding to polyatomic molecules. *Nature* **606**, 688–693 (2022). <https://doi.org/10.1038/s41586-022-04703-3>

- Received: 08 July 2021
- Accepted: 30 March 2022
- Published: 22 June 2022
- Issue Date: 23 June 2022
- DOI: <https://doi.org/10.1038/s41586-022-04703-3>

## Share this article

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

[Get shareable link](#)

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

[Copy to clipboard](#)

Provided by the Springer Nature SharedIt content-sharing initiative

---

This article was downloaded by **calibre** from <https://www.nature.com/articles/s41586-022-04703-3>

- Article
- Open Access
- [Published: 22 June 2022](#)

# Engineering topological states in atom-based semiconductor quantum dots

- [M. Kiczynski](#),
- [S. K. Gorman](#),
- [H. Geng](#),
- [M. B. Donnelly](#),
- [Y. Chung](#),
- [Y. He](#),
- [J. G. Keizer](#) &
- [M. Y. Simmons](#)

*Nature* volume **606**, pages 694–699 (2022)

- 1271 Accesses
- 179 Altmetric
- [Metrics details](#)

## Abstract

The realization of controllable fermionic quantum systems via quantum simulation is instrumental for exploring many of the most intriguing effects in condensed-matter physics<sup>1,2,3</sup>. Semiconductor quantum dots are particularly promising for quantum simulation as they can be engineered to achieve strong quantum correlations. However, although simulation of the

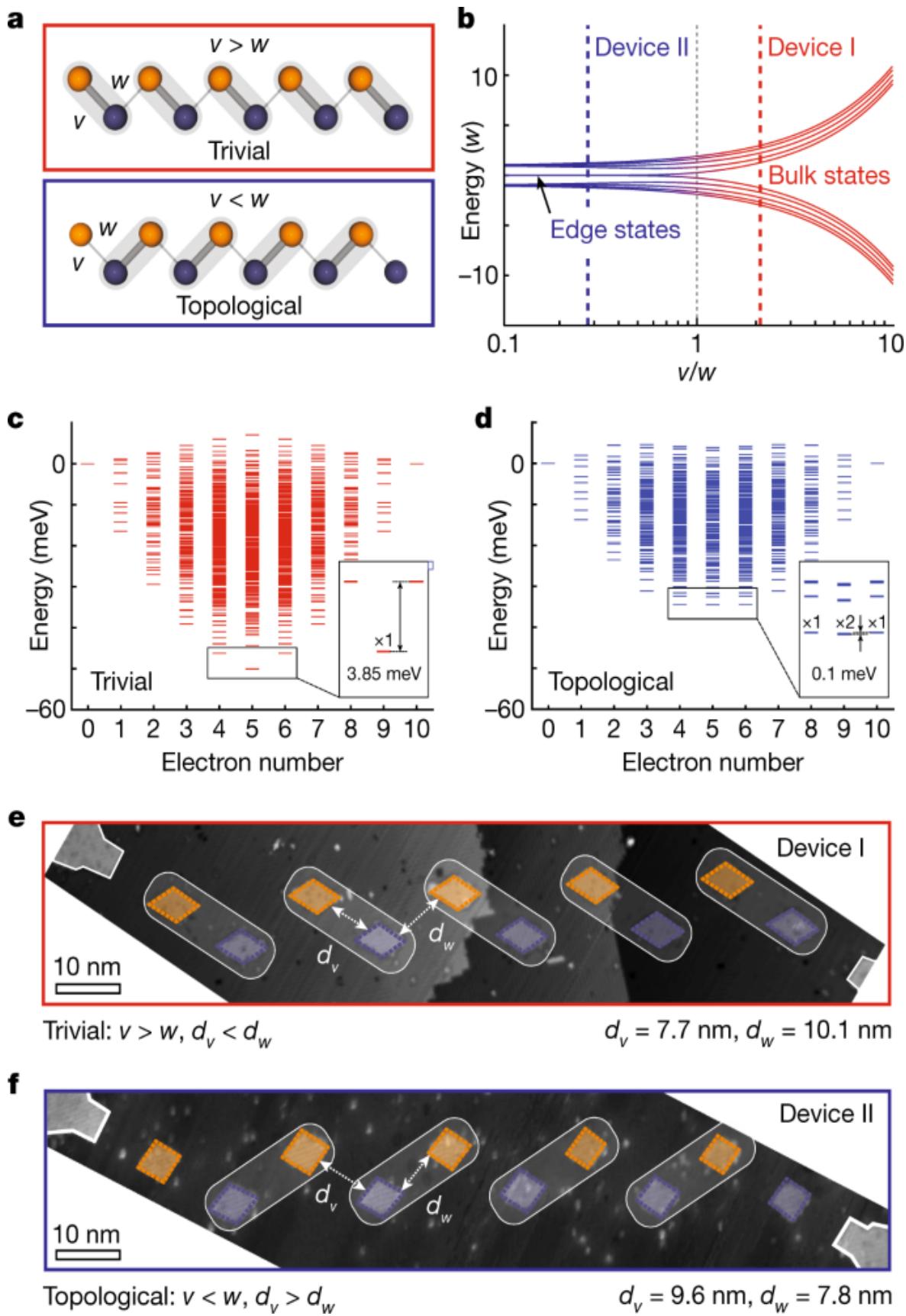
Fermi–Hubbard model<sup>14</sup> and Nagaoka ferromagnetism<sup>5</sup> have been reported before, the simplest one-dimensional model of strongly correlated topological matter, the many-body Su–Schrieffer–Heeger (SSH) model<sup>6,7,8,9,10,11</sup>, has so far remained elusive—mostly owing to the challenge of precisely engineering long-range interactions between electrons to reproduce the chosen Hamiltonian. Here we show that for precision-placed atoms in silicon with strong Coulomb confinement, we can engineer a minimum of six all-epitaxial in-plane gates to tune the energy levels across a linear array of ten quantum dots to realize both the trivial and the topological phases of the many-body SSH model. The strong on-site energies (about 25 millielectronvolts) and the ability to engineer gates with subnanometre precision in a unique staggered design allow us to tune the ratio between intercell and intracell electron transport to observe clear signatures of a topological phase with two conductance peaks at quarter-filling, compared with the ten conductance peaks of the trivial phase. The demonstration of the SSH model in a fermionic system isomorphic to qubits showcases our highly controllable quantum system and its usefulness for future simulations of strongly interacting electrons.

## Main

Superconductivity, magnetism<sup>12</sup>, low-dimensional electron transport<sup>13</sup>, topological phases<sup>14</sup> and other exotic phases of matter arise owing to the presence of strongly interacting particles within crystals<sup>15</sup>. However, the complexity of simulating such large quantum systems becomes intractable using classical computing methods<sup>16</sup>. A promising solution is to build a physical system at the same scale so that we can simulate these interacting fermionic systems<sup>17,18</sup> directly, known as analogue quantum simulation<sup>19,20</sup>. The Su–Schrieffer–Heeger (SSH) model is the prototypical example of topological matter that describes a single electron hopping along a one-dimensional dimerized lattice with staggered tunnel couplings,  $v$  and  $w$ , as shown in Fig. 1a<sup>21</sup>. The SSH model has been experimentally simulated in physical systems of varying dimensions from Rydberg atoms (about 10 μm) to mechanical systems (about 10 mm) (Table 1). The coupling strengths of the various simulators lie in the nanoelectronvolt to microelectronvolt range, limiting their ability to reach the fully coherent regime. Importantly, these

systems can be readily solved classically as they do not simulate many-body interactions. Only recently has the interacting many-body SSH model been observed using Rydberg atoms with an effective infinite on-site interaction (hardcore bosons)<sup>[10](#)</sup>. The ability to control the interaction strength, however, is critical for investigating fermionic systems<sup>[22](#) [23](#)</sup>.

**Fig. 1: Realization of the SSH model in precision-engineered phosphorus donors in silicon.**



**a**, A dimerized one-dimensional lattice with staggered hopping amplitudes (tunnel couplings)  $v$  and  $w$ . The array consists of two sublattices where there exists a trivial phase,  $v > w$  (top) and a topological phase  $v < w$  (bottom). **b**, Single-particle energy spectrum of the SSH model for a linear array of ten quantum dots as a function of the interdot coupling ratio  $v/w$ . For  $v/w < 1$ , there exists zero-energy topological edge states, whereas the trivial case  $v/w > 1$  exhibits an excitation gap. **c, d**, Calculated multi-electron energy spectrum in the trivial (**c**) and topological (**d**) phases for different electron numbers at quarter-filling. The trivial array exhibits a single ground state with five electrons about 3.85 meV below the four- and six-electron states, whereas the topological phase exhibits a nearly four-fold degeneracy involving four, five (two-fold degenerate) and six electrons. **e, f**, Scanning tunnelling micrograph for the trivial (**e**; device I) and topological (**f**; device II) phases. The lighter regions show the open lithographic hydrogen mask. The devices consist of an array of  $N = 10$  Coulomb-confined quantum dots with staggered nearest-neighbour distances, tunnel-coupled to a source (drain) lead at the start (end) of the array outlined in white to perform bias spectroscopy. Device I is designed to be in the trivial phase with  $d_v = 7.7 \pm 0.1$  nm and  $d_w = 10.1 \pm 0.2$  nm, and device II is designed to be in the topological phase with  $d_v = 9.6 \pm 0.4$  nm and  $d_w = 7.8 \pm 0.6$  nm.

### Table 1 Experimental demonstrations of the SSH model to date

Semiconductor quantum dots are an emerging platform for the quantum simulation of strongly correlated electron systems<sup>4,5,20</sup>. Interacting electrons confined to quantum dots have been described by the Hubbard model<sup>16</sup> involving Coulombic interactions that describe the energy required to add electrons to the same (on-site,  $U$ ) or neighbouring (intersite,  $V$ ) quantum dot<sup>24</sup>. Here intersite hopping is governed by the tunnel coupling,  $t$ , between quantum dots, and each dot can be tuned using electrostatic gates to raise or lower their energy levels,  $\langle \{ \epsilon \} \rangle$  (ref. <sup>4</sup>). Phosphorus donors in silicon in particular have been proposed as promising candidates for simulators as they are nanoscale in size with very strong on-site energies ( $U \approx 25$  meV) and can be engineered to have strong intersite ( $V \approx 5$  meV) and hopping ( $t \approx 5$  meV) energies, while operating with a low thermal energy of  $k_B T \approx 0.02$  meV (ref. <sup>25</sup>), where  $k_B$  is the Boltzmann constant and  $T$  is the temperature, reaching a range of  $U/t \approx 1\text{--}100$  while remaining in the low-

temperature limit with  $t/k_B T > 10$  (ref. [23](#)). The ability to reach the low-temperature, strongly interacting regime allows for a number of coveted quantum phases, such as superconductivity<sup>[26](#)</sup> and antiferromagnetism<sup>[27](#)</sup>, to be simulated<sup>[28](#)</sup>. Despite the promise of semiconductor simulators, significant challenges have remained to simulate full quantum systems. These relate to the ability to precisely engineer, and tune, both the large on-site interaction energies and tunnel couplings to allow for the formation of a well defined coherent state across the system. In particular, for 10 quantum dots, we require precision control across 110 different experimental parameters related to  $U, V, t$  and  $\{\epsilon\}$ .

In this paper, we utilize the atomic-precision placement accuracy of the scanning tunnelling microscope (STM) to engineer quantum dots with large on-site energies ( $U \approx 25$  meV) and uniform size to realize a homogeneous linear array for reliable simulation accuracy. If the quantum dots are too big, the capacitive coupling between individual dots becomes too large to independently control them. Conversely, if they are too small, then a small change in the number of phosphorus donors within the quantum dot can substantially change the on-site energy, leading to randomness in the array. Importantly, our subnanometre-precision capability allows us to change the values of  $v$  and  $w$  with millielectronvolt resolution so that we can reliably enter both topologically trivial and topologically non-trivial regimes. Finally, a substantial challenge for gate-defined quantum dot architectures is that they require electrostatic gates to create the quantum dot potential and control the tunnel couplings requiring a minimum of about two gates per quantum dot<sup>[29,30,31,32](#)</sup>. With donor-based dots, we do not require these additional confinement gates and require only six electrostatic gates to control a ten-quantum dot array, thereby avoiding unnecessary cross-talk between gates. To ensure the creation of a well defined quantum state across the array, we designed an iterative maximum-current-alignment procedure to align the quantum dot energy levels within approximately 0.5 meV. The quantum state formed is then measured using bias spectroscopy via the planar source and drain leads. Having determined the necessary conditions to form the desired state, we simulate the one-dimensional topological phases associated with the interacting SSH model<sup>[21](#)</sup>.

The SSH model is one of the simplest known instances of topological quantum systems. The eigenenergies (Fig. 1b) of the SSH model give rise to two distinct phases dependent on the ratio of tunnel couplings with a topological phase transition at  $v=w$ . For  $v>w$ , there is a topologically trivial phase, where the lattice acts as a bulk insulator with the electron delocalized across the array and an energy gap between the upper and lower bulk states. For  $v<w$ , there exists a topologically non-trivial phase—a symmetry-protected topological phase—which gives rise to two zero-energy edge states where the electron is localized at the edge sites of the lattice<sup>33</sup>.

In previously measured instances of the SSH model, particle–hole symmetry was conserved owing to a lack of intersite electron–electron interactions<sup>21</sup>. However, quantum dots in semiconductors are affected by the intersite Coulomb interaction,  $V_{i,j}$  which is the change in energy of quantum dot  $j$  owing to the addition of an electron on quantum dot  $i$ . These long-range electron–electron interactions break the particle–hole symmetry, leading to non-degenerate electron and hole states<sup>4</sup>. As a consequence, it is important to control both the electron filling of the array and changes in the electrostatic environment to ensure correct simulation results. This requires us not only to have independent control of each quantum dot potential but also to alter the energy levels in unison—a tremendous technological challenge for such a small array with strong tunnel couplings and on-site energies ([Supplementary Section II](#)). As a result, we consider the full Hamiltonian of the extended (spinless) Hubbard model for a linear array of  $N$  quantum dots, given by

$$\begin{aligned} \text{\$\$}\{H\}_{U}=&\mathop{\sum }\limits_{i=1}^N\{\{\epsilon\}_{i}\\ n_i+\mathop{\sum }\limits_{i}^N\{U\}_i\}{n_i}\\ (&n_i-1)+\mathop{\sum }\limits_{i}^{N-1}\{t\}_{i,i+1}\\ (&c)_i^{\dagger}\{c\}_{i+1}+\{\rm h\}\{\rm c\}).+\mathop{\sum }\limits_{i,j}^N\{V\}_{i,j}\{n\}_i\{n\}_j,\$\\ (1) \end{aligned}$$

where  $\{\{\epsilon\}_i\}$  is the energy levels of the  $i$ th dot of the array,  $n_i$  is the dot occupation operator,  $t_{i,i+1}$  is the tunnel coupling between nearest-neighbour  $i$ th and  $i+1$  dots,  $U_i$  is the on-site Coulomb interaction term,  $V_{i,j}$  is the intersite Coulomb interaction terms between the  $i$ th and  $j$ th dots,

$(\{c\}_i^{\dagger})$  are the creation (annihilation) operator of a particle on site  $i$  and h.c. indicates Hermitian conjugate.

In Fig. 1c,d, we show the 1,024 calculated multi-electron energy levels based on the two experimental arrays fabricated in the trivial and topological phases respectively, ordered by the number of electrons in the quantum dot array at quarter-filling. For ten quantum dots, in which each dot can host two electrons, quarter-filling corresponds to where there are five electrons shared across the dots. Here, if the array is in the trivial phase, the electrons will simply arrange themselves into each dimer such that they are spread evenly across the array. In contrast, if the array is in the topological configuration at quarter-filling, four electrons arrange themselves in the middle of the array (similar to the trivial phase). However, the fifth electron cannot occupy a dimer. We focus on quarter-filling of the array as the interacting topological states involve this fifth electron becoming localized to each end of the array. In the trivial phase ( $v > w$ ), there is a singly degenerate ground state (box in Fig. 1c) delocalized across the entire array with a large energy gap of about 3.85 meV separating the five-electron state from the four- and six-electron states. The topological phase ( $v < w$ ) exhibits a nearly four-fold degenerate (about 0.1 meV) ground state (box in Fig. 1d) involving four, five (two-fold degenerate) and six electrons. The ground state corresponds to four electrons on the inner eight quantum dots of the array with either zero, one or two additional electrons localized at the two edge quantum dots.

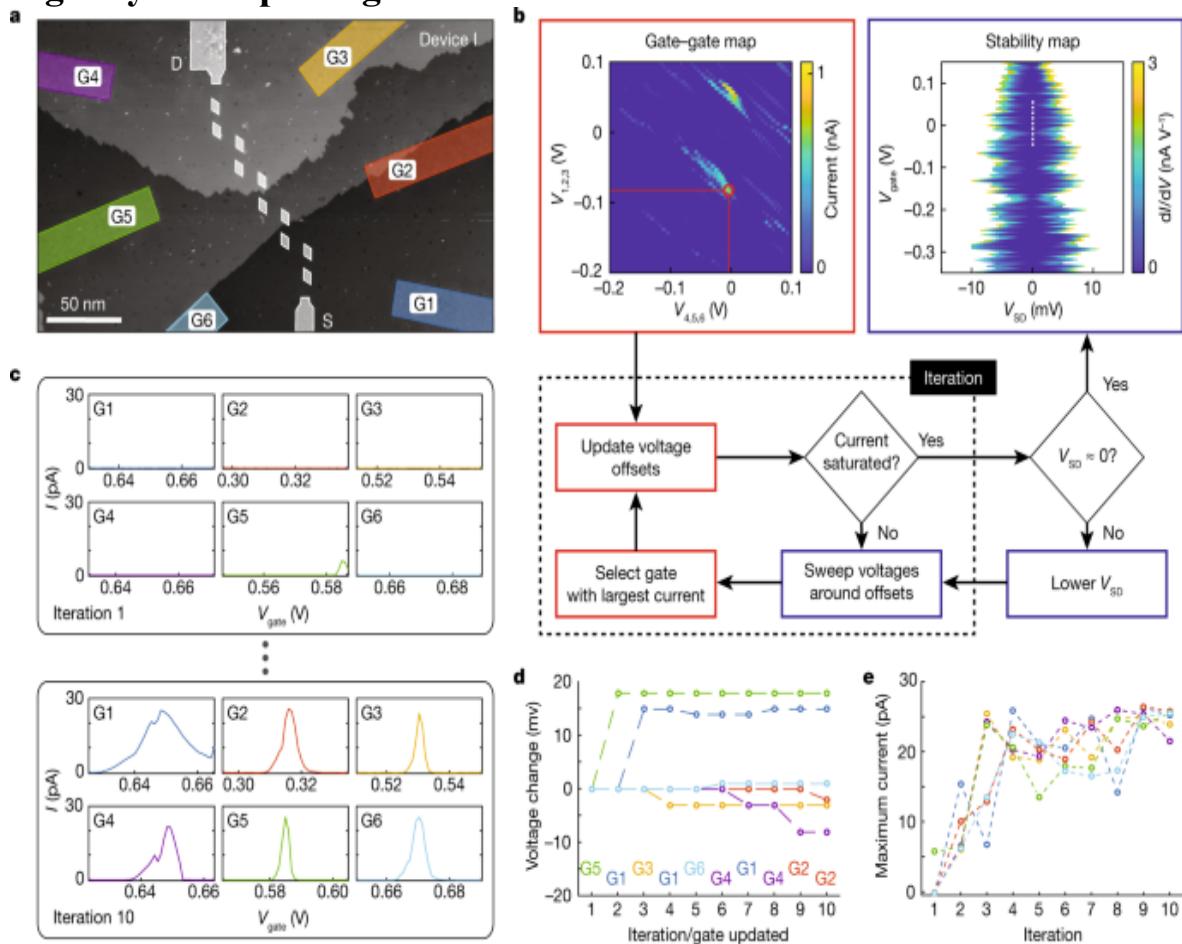
The experimental realization of two such arrays is shown in Fig. 1e,f, where the quantum dots are tunnel-coupled to source and drain leads to perform bias spectroscopy through the array. The regions outlined in orange and blue are the quantum dots, and the two regions at the start and end of the array outlined in white are the source and drain leads. The tunnel couplings  $t_{i,i+1}$  are engineered via the interdot separation,  $d_{i,i+1}$  and follow an inverse exponential dependence,  $\propto \exp(-2d_{i,i+1}/3)$  (ref. 34). By staggering the quantum dot array, we have ensured that non-nearest-neighbour tunnelling is exponentially suppressed with an estimated  $t_{i,i+2}/t_{i,i+1} \approx 0.01$ , ensuring that electron transport occurs in series through the array, while maximizing the differential lever arms to the dots (Supplementary Section I). The SSH model requires that the tunnel couplings have alternating strengths to observe the different topological

phases while being simultaneously large enough to allow for a measurable transport current for bias spectroscopy. The quantum dot size is critical as the confinement potential experienced by the last outer electron, and hence the wavefunction overlap to the neighbouring quantum dots, depends on the number of donors comprising the quantum dot. We therefore fabricate quantum dots with an area of about  $25 \text{ nm}^2$  (about 25 phosphorus donors per site<sup>35</sup>). The nanoscale size of the quantum dots allows us to achieve large on-site energies of about 25 meV, but, importantly, where a small change in the size of the quantum dot does not significantly alter  $U$ ,  $V$ , or  $t$ . The small separation of  $\leq 10 \text{ nm}$  also allows us to achieve large tunnel couplings  $t \approx 1\text{--}10 \text{ meV}$  for these quantum dot sizes<sup>35</sup>. Device I (Fig. 1e) is designed to be in the trivial phase with average staggered quantum dot distances,  $d_v = 7.7 \pm 0.1 \text{ nm}$  and  $d_w = 10.1 \pm 0.2 \text{ nm}$ , corresponding to  $\langle v/w \rangle = 2.08$ . Device II (Fig. 1f) is designed to be in the topological phase where we now engineer an average staggered donor distance,  $d_v = 9.6 \pm 0.4 \text{ nm}$  and  $d_w = 7.8 \pm 0.6 \text{ nm}$ , corresponding to  $\langle v/w \rangle = 0.265$ . These values highlight the subnanoscale accuracy that we can engineer devices with using STM lithography so that we can change  $\langle v/w \rangle$  between 0.265 and 2.08 (ref. 36).

Figure 2a shows an STM image of the full device I. Here the outlined lighter regions show the lithographic hydrogen mask with six capacitively coupled control gates (G1 to G6), crucial to independently control the energy levels of the quantum dots (device II gate structure is nominally identical). Owing to the unique geometry of the device, the total lever arms of all gates linked together to each quantum dot are engineered to be consistent with a variation of less than 2.5%. This small variation means that we can also raise the global energy level of the whole quantum dot array for bias spectroscopy to measure the different phases of the SSH model. To align the energy levels of the quantum dots, we used a maximum-current-alignment scheme, in which the individual gates are tuned as outlined in Fig. 2b. This is achieved by initially setting the gate voltages at a conductance peak determined by sweeping G1, G2 and G3, against G4, G5 and G6, while measuring the current through the array. While positioned at this conductance peak, each gate is then individually swept around a set value, whereas all other gates were kept constant, as illustrated in Fig. 2c. After sweeping all six gates about their set voltage, the largest current peak is found and the

corresponding gate is updated to the voltage at the centre of the current peak (G5 in the first iteration shown in Fig. 2d). All gates are then swept again, repeating this process, updating a single gate at a time as shown in Fig. 2d. Figure 2e shows the maximum current measured on each gate sweep per iteration for a constant source–drain bias,  $V_{SD}$ . When the maximum current plateaus,  $V_{SD}$  is reduced further and the entire process is repeated to increase the alignment accuracy.

**Fig. 2: Maximum-current-alignment scheme of the quantum dot array using only six in-plane gates.**

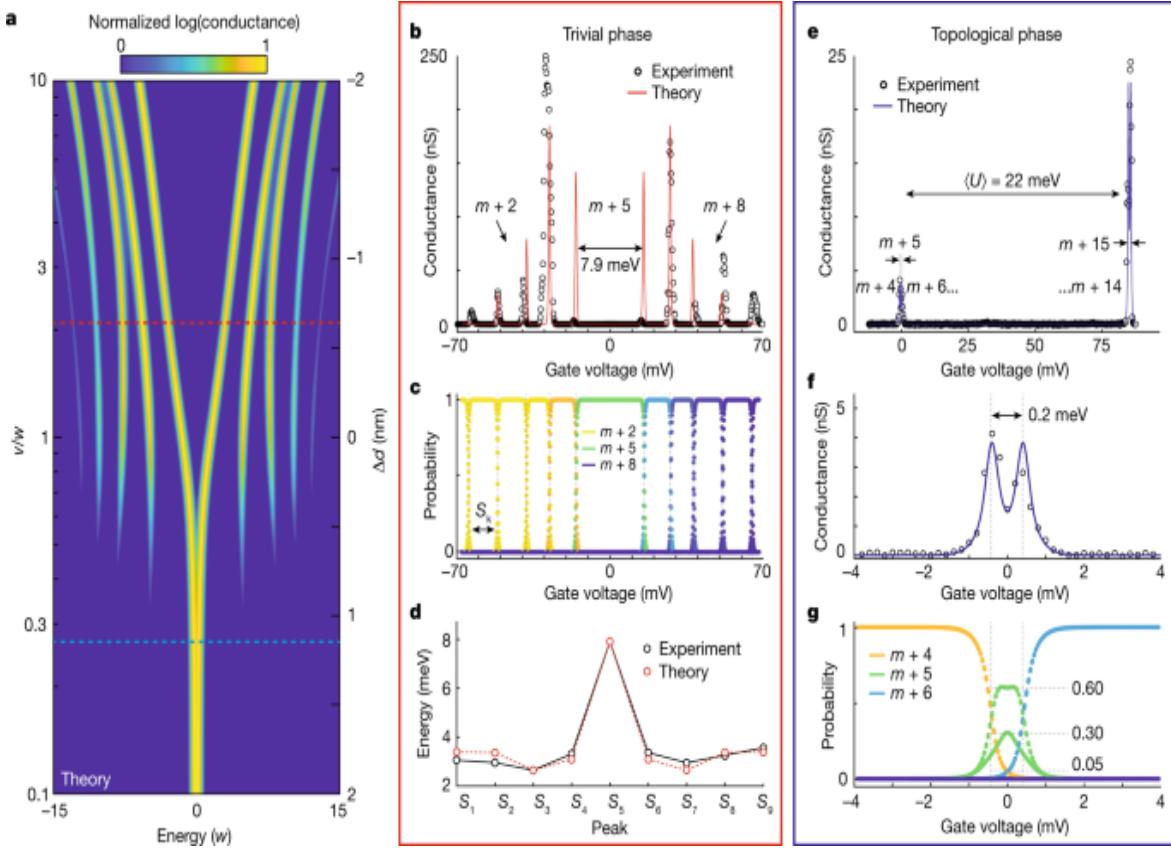


**a**, An STM micrograph of device I, which shows the six control gates, labelled G1 to G6, used to tune the energy levels of the quantum dots to observe the conductance peaks using bias spectroscopy. Here we can observe the atomic step height of the silicon surface and the nanoscale size of the array. **b**, Schematic of the protocol used to align the quantum dots in the array. The quantum dots can be brought into alignment by varying the

voltages applied to the control gates to tune the quantum dots for maximum current through the array. Each control gate is initially set to a specific value, chosen from a conductive region from the current map (red circle) while changing gates G1–G3 and G4–G6 together. The voltage on each gate is then swept, in turn, around their respective maximum current values, while the other gates are kept constant. After all gate voltages have been swept, a single gate value is then updated corresponding to the maximum current measured. The process is then repeated updating one gate each time. When the maximum current saturates, the source–drain bias,  $V_{SD}$ , is then reduced and the control gates are retuned again to increase the maximum current. Once the  $V_{SD}$  is near zero, a stability diagram is measured as shown in the top right and the zero-bias conductance (dotted white line) is used for comparison with the simulated SSH model in Fig. 3. **c**, Examples of the individual gate sweeps on the first iteration (top) and on the tenth iteration (bottom) for a constant  $V_{SD}$ . **d, e**, The voltage on each gate per iteration (the gate updated per iteration is labelled at the bottom; **d**) and showing the maximum current measured on each gate sweep per iteration (**e**).

Once the energy levels of the quantum dots are aligned, we perform a stability diagram measurement by shifting the energy levels of all quantum dots to investigate the electron occupation of the array. The stability diagram allows us to determine the electron occupation of the array as a function of the energy levels of the quantum dots. At zero source–drain bias (dashed white line in Fig. 2b), there is only enough energy to add a single electron to the array at a time. In this regime, we simulate the SSH model in the two devices. Figure 3 shows a comparison of the experimental and theoretical results of the two devices. Figure 3a shows the calculated normalized zero-bias conductance as a function of the ratio of tunnel couplings ( $v/w$ ) of the array. The design of device I in the trivial phase is given by the dashed red line  $\langle v/w \rangle = 2.08$ , whereas device II in the topological phase is given by the dashed blue line  $\langle v/w \rangle = 0.265$ .

**Fig. 3: Experimental signature of the SSH model in precision-engineered quantum dot arrays.**



**a**, A theoretical map of the normalized log(conductance) as a function of the ratio of tunnel couplings (with the intersite Coulomb interactions given by a  $1/d^{1.5}$  dependence, where  $d$  is the quantum dot separation), with device I given by the dashed red line and device II given by the dashed blue line. **b**, Conductance trace obtained at zero source–drain bias ( $V_{SD} = 0$ ), while shifting the energy levels of all quantum dots, in the trivial phase. We observe 10 conductance peaks corresponding to a single Hubbard band with a gap of about 7.9 meV (about  $2 \times 3.85$  meV) at quarter-filling. **c**, The occupation probability of the many-body eigenenergies (all 1,024) of the Hubbard model as a function of the combined gate voltage for the trivial phase. The conductance peaks in **b** correspond to transitions between the different electron number ground states (grey dashed lines), which are separated in gate voltage by  $S_k$ . **d**, We use the extracted values of  $S_k$  from the experimental results to obtain the parameters from the Hubbard model (Methods) and compare them to theory. **e**, In the topological phase, we observe two close conductance peaks around zero gate voltage resulting from the topological edge states at quarter-filling, whereas no conductance peaks are observed away from quarter-filling (between 0 mV and 85 mV on

the gates). The conductance peaks (there are two closely spaced) observed around 85 mV correspond to the addition of 10 extra electrons to the array and are separated by  $\langle U \rangle$ , which is the average on-site energies of the array. **f**, A zoom-in of the two conductance peaks of the topological phase corresponding to the  $m + 4 \rightarrow m + 5$  and  $m + 5 \rightarrow m + 6$  electron transitions to the array shown at 0 mV in **e**. **g**, The occupation probability of the topological phase as a function of the combined gate voltage. There is a sharp transition from  $m + 4$  electrons where no electrons exist on the two edge sites of the array to  $m + 6$  electrons in which the edge sites of the array are fully occupied. At quarter-filling ( $m + 5$  electrons), there is a non-zero probability that there is either 4 ( $P \approx 0.05$ ), 5 ( $P \approx 0.30$  and  $P \approx 0.60$  owing to the two-fold degeneracy) or 6 ( $P \approx 0.05$ ) electrons existing on the array as a result of the nearly four-fold degenerate ground state of the topological phase. This remarkable feature of the two conductance peaks separated by about 0.2 meV is the signature of the topological phase of the many-body SSH model.

Figure 3b shows the zero-bias conductance as a function of the combined voltages on all the gates obtained from the trivial phase (device I) with the theoretical calculation shown in red. There are ten conductance peaks corresponding to a change in the total number of electrons on the array (Fig. 3c). We control the electron filling of the array by adjusting the gate voltages to tune the electron number from  $m$  to  $m + 10$  (half-filling). At a quarter-filling ( $m + 5$ ), there is a gap in the energy spectrum (7.9 meV, approximately twice the gap of 3.85 meV in Fig. 1c) corresponding to the single ground state of the SSH model for the trivial phase (Fig. 1c). From the estimate of  $V_{i,j}$  from electrostatic modelling, and fitting the magnitude of the tunnel coupling (Methods), we model the array to obtain the width of the different electron number regions,  $S_k$  (width of the  $m + k \rightarrow m + k + 1$  region). Figure 3d shows the width of the experimentally measured stability regions, obtained by determining the energy between adjacent conductance peaks, compared with the theoretical calculations based on electrostatic modelling with a tunnel coupling ratio  $\langle v/w \rangle = 2.08$ . We find excellent agreement between the experimental and theoretical values. Small discrepancies ( $\lesssim 1$  meV) are most likely due to small misalignments of the quantum dot energy levels, which gives rise to on-site

disorder, causing small shifts in the conductance peaks such that the peak structure is no longer symmetric around zero ([Supplementary Section III](#)).

We now look at the topological phase of the SSH model (device II) shown in Fig. [3e](#). The blue line in Fig. [3e](#) represents the theoretical fit to the experiment, with a tunnel coupling ratio ( $\langle v/w \rangle = 0.265$ ). We show a similar voltage range scan as for device I but here we observe only two sets of closely spaced peaks at zero gate voltage and at 85.5 mV corresponding to the average on-site energies across the array, ( $\langle U \rangle = 22.0 \pm 3.2$  meV). The conductance peaks from the states away from quarter-filling are not visible as they are now delocalized within the bulk of the array with a low probability of existing at the edge quantum dots. As a result, tunnelling between these bulk-like states and the source and drain leads is significantly suppressed. In the topological phase, the quarter-filling gap almost disappears completely with a sharp transition from the  $m + 4$  to  $m + 6$  states given by only two conductance peaks separated by about 0.2 meV, shown in Fig. [3f](#). These electron states correspond to where there are no electrons ( $m + 4$  electrons) on the edge quantum dots to where both dots are occupied ( $m + 6$  electrons). Importantly at exactly quarter-filling ( $m + 5$  electrons), there is a non-zero probability that either four ( $P \approx 0.05$ ), five ( $P \approx 0.30$  or  $P \approx 0.60$  owing to the two-fold degeneracy) or six ( $P \approx 0.05$ ) electrons exist on the array at the same time owing to the nearly four-fold-degenerate ground state, as illustrated in Fig. [3g](#). This remarkable observation, that at zero gate bias there is a superposition of the number of electrons on the edge quantum dots, is a result of the near-zero energy of the topological states of the array and is a distinctive property of the many-body SSH model. As these topological states are localized at the edge quantum dots, the current flowing through the array corresponds to an electron moving from one side of the array to the other without occupying any of the inner quantum dots. This unique property is a direct consequence of the topology embedded within the SSH model as confirmed by the double conductance peak in Fig. [3f](#).

In conclusion, we have observed clear signatures of the topological states of the interacting SSH model in semiconductor quantum dots. To achieve this, we have precision engineered two devices with subnanometre resolution consisting of a linear array of ten donor-based quantum dots, with staggered nearest-neighbour tunnel couplings. The minimal gate design of our

epitaxial devices allows for both individual and global alignment of all the energy levels of the quantum dots such that bias spectroscopy can probe the topological phases of the array. Importantly, we confirm the existence of the one-dimensional topological phase at quarter-filling ( $\langle v/w \rangle = 0.265$ ) with the observation of two overlapping peaks in the zero-bias conductance corresponding to the near four-fold degeneracy of the many-body SSH model. For the trivial phase ( $\langle v/w \rangle = 2.08$ ), we observe ten zero-bias conductance peaks corresponding to delocalized states across the entire array with an energy gap around quarter-filling owing to the large interdot coupling of about 3.5 meV. The low-gate-density design and low noise<sup>37</sup> exhibited in these all-epitaxial devices offers a promising platform for simulating strongly interacting electron systems for quantum chemistry applications<sup>19</sup>. Future work will focus on extending the size of the quantum dot arrays, incorporating charge sensors and extending the simulations to engineered two-dimensional lattices<sup>38</sup>.

## Methods

### Device fabrication

The devices were fabricated on a 1–10  $\Omega$  cm p-type natural silicon substrate. The substrate was prepared via a series of high-temperature anneals, up to 1,100 °C, followed by a controlled cooling to 330 °C, resulting in a (2 × 1) reconstructed surface. The substrate was then terminated with atomic hydrogen, which can be selectively desorbed using the STM tip, leaving a hydrogen lithographic mask representing the device. After STM lithography, the substrate was exposed to a phosphine (PH<sub>3</sub>) precursor gas, in which phosphorus was absorbed and incorporated, at 350 °C, into the exposed areas. The device was then encapsulated with 40 nm of natural silicon using molecular beam epitaxy at a rate of about 0.125 nm min<sup>-1</sup>. A more detailed description of STM hydrogen lithography and the further device processing to electrically contact the device can be found in refs. <sup>42,43,44</sup>.

### Experimental measurement setup

The electrical measurements were performed at millikelvin temperatures inside a dilution refrigerator. The two devices were measured in different fridges. Device I was measured in a fridge with a base temperature of about 10 mK, and device II was measured in a fridge with a base temperature of about 100 mK. Extended Data Fig. 1 shows a schematic of the electrical connections to the device. To perform the measurements, voltage sources were connected to the source and gates, to control their chemical potentials, and the transport current was measured through the drain. The drain current was amplified and converted to a voltage signal using a FEMTO variable-gain low-noise current amplifier DLPICA-200 with a low-pass 10-Hz filter. The filter signal was then digitized using a National Instrument data acquisition box (NIDAQ). The voltage sources used for the two devices were different. For device I, the source voltage was generated by the auxiliary of a Stanford Research Systems SR830 DSP lock-in amplifier and 1/50 room-temperature resistive-voltage divider, whereas the gate voltages were generated by two NIDAQs. Device II was measured on a different dilution refrigerator and had a different experimental setup. The source voltage was generated by a NIDAQ and a 1/50 room-temperature resistive-voltage divider. Gates 1, 2 and 3 voltages were also generated by the NIDAQ, whereas gates 4, 5 and 6 voltages were each generated by a Yokogawa 7651 programmable d.c. source. The different voltage sources used to measure the different devices should not affect the measurement results in a fundamental way.

## Calculation of the parameters of the extended Hubbard model

The quantum dot array can be described by the extended Hubbard model with the Hamiltonian given in equation (1). To theoretically solve the extended Hubbard model and calculate the parameters of the array, the data were fit using an open-source python package QmeQ<sup>45</sup>. The quantum dot Hamiltonian is described by the single-particle states  $\langle \{\sum_{i=1}^N \{\epsilon_i\} n_i + \sum_{i=1}^{N-1} t_{i,i+1} c_i^\dagger c_{i+1} + \{rm{h}\} \{rm{c}\} \rangle$  and the Coulomb matrix elements,  $\langle \sum_{i,j}^N V_{i,j} n_i n_j \rangle$ . The Hamiltonian is constructed in the Fock basis<sup>45,46</sup>, for example,  $\langle |0010101110\rangle$ , where 1 indicates an electron at that quantum dot and 0 indicates the quantum dot is unoccupied, and diagonalized exactly to obtain the exact

many-body eigenstates  $\langle |a\rangle \rangle$ ,  $\langle H = \sum \{E\}_a |a\rangle \rangle$ . The transport current through the quantum dot array is then calculated numerically using the Pauli master equation. The array is assumed to be weakly coupled to the source and drain leads at a temperature,  $T$ , with a density of states that follows a Fermi distribution,  $f(E)$ .

Analogous to the non-interacting energy-level diagram shown in Fig. 1b, we also calculated the energy-level diagram of the interacting system. As with Fig. 3a, we assume that the intersite Coulomb terms follow a  $d^{-1.5}$  dependence<sup>47</sup>, and that the energy levels of the quantum dots are tuned via  $\{\{\epsilon_i\}\} = -\{\sum_j N_j V_{ij}\}$ . Extended Data Fig. 2 shows the energy spectrum for an array of ten sites for the  $m+4$ ,  $m+5$  and  $m+6$  electron states at quarter-filling, showing the more complex excited-state spectrum of the many-body states of the array.

For  $v/w > 1$ , at quarter-filling ( $m+5$  electrons), there is a large gap between the ground  $m+5$  and the lowest  $m+4$  and  $m+6$  electron states with a single ground state (labelled ‘bulk states’ analogous to Fig. 1b). By changing the tunnel couplings such that  $v/w < 1$ , the ground state becomes doubly degenerate with  $m+5$  electrons (labelled ‘edge states’ analogous to Fig. 1b). Here the quarter-filling gap is greatly reduced resulting in nearly degenerate states with differing electron numbers. The small energy gap at  $v/w < 1$ , is observed in the conductance trace in Fig. 3f and reflects the intersite Coulomb interactions present in the system.

In Fig. 3b and Fig. 3e, we show the theoretical calculation of the conductance through the quantum dot array fitted to the experimental data for the trivial and topological phases, respectively. In the theoretical calculation, we consider a spinless ten-dot array in the regime where there can be at most one electron on each of the quantum dots. Electron transport is restricted to sequential tunnelling through the array, as a result of the engineering of the device design, and only a single electron can tunnel through the quantum dots. No higher-order co-tunnelling events are allowed, as the source–drain bias is sufficiently small, making the choice of the Pauli master equation valid. All intersite Coulomb interaction terms,  $V_{i,j}$  are included and accounted for by tuning the energy levels of the dots such that  $\{\{\epsilon_i\}\} = -\{\sum_j N_j V_{ij}\}$ .

The tunnel coupling and intersite Coulomb interaction parameters of the extended Hubbard model are obtained by fitting to the measured conductance peaks in the trivial phase, and using the measured distances from the STM and the electrostatic modelling. The trivial phase was used for the fit, as in this phase all ten conductance peaks can be observed, whereas for the topological phase only two conductance peaks are observed. The conductance peaks in the trivial phase in Fig. 3b each correspond to transitions between different particle number ground states, which are separated in gate voltage by  $S_k$  (the voltage separation between the  $k$  and  $k+1$  particle number). These peak separations,  $S_k$ , are dominated by the tunnel coupling strengths,  $v$  and  $w$ , and the intersite Coulomb interaction strengths,  $V_{i,j}$ . Using the extracted values of  $S_k$  from the experimental results, we fit the overall magnitude of the tunnel coupling and intersite Coulomb interactions (while keeping the theoretically determined trends as a function of distance) to find the parameters shown in [Supplementary Tables 4–6](#).

The tunnel coupling,  $t$ , is engineered via the interdot donor separation,  $d$ , and follows an exponential dependence,  $t = t_M \exp(-2d/3)$  eV, where the tunnel coupling magnitude  $t_M = 0.1742$  for a 1P–2P quantum dot system<sup>34</sup>. We anticipate a similar dependence of  $t$  as a function of interdot separation distance; however,  $t_M$  is now used as a fitting parameter as it is known to depend on the crystallographic orientation of the quantum dots with  $v$  (about  $\langle 100 \rangle$ ) and  $w$  (about  $\langle 120 \rangle$ ) having different scaling factors owing to their different interdot axes. [Supplementary Table 6](#) shows the distances and tunnel couplings for the two devices, with  $\langle v/w \rangle = 2.08$  (1.702, 2.460) for the trivial phase and  $\langle v/w \rangle = 0.265$  (0.146, 0.539) for the topological phase. The best fit to the zero-bias conductance peaks are achieved when the  $w$  tunnel couplings are approximately twice the  $v$  couplings for the same distance ( $t_{M,w} \approx 2t_{M,v}$ ). This angular dependence on the tunnel coupling between individual donors is a well known consequence of the silicon crystal lattice<sup>48,49,50</sup>; however, it has now been observed directly for quantum dots of larger size. This additional knowledge will assist in future experimental designs for finer engineering of the tunnel couplings.

The extracted tunnel-coupling parameters give excellent fits to the experimental data as seen in Fig. 3b,e, with Fig. 3d showing a comparison of the peak separations,  $S_k$ , for the experimental and theoretical data. Small variations between the experimental and theoretical data can be attributed to errors in the electrostatic modelling and small offsets in the alignments of the quantum dots.

The presence of intersite Coulomb interactions gives rise to an approximately 0.2 meV splitting observed in the topological states as seen in Fig. 3f. Extended Data Fig. 3 shows the calculated conductance for the topological phase for varying intersite Coulomb interaction strengths,  $V_{i,j}$  with values given in [Supplementary Tables 5, 6](#). With no intersite Coulomb interactions, a single conductance peak is observed (Extended Data Fig. 3a), whereas the inclusion of intersite Coulomb interactions gives rise to two conductance peaks, in which the separation between the peaks increases with increasing  $V_{i,j}$ . In addition to the presence of intersite Coulomb interactions, a small splitting (about 0.04 meV) in the topological states would arise owing to the non-zero  $v/w$  ratio and the finite length of the chain. Here the topological states, exponentially localized at opposite ends of the chain, have a finite overlap. This small splitting, as presented in Extended Data Fig. 3a, is, however, too small to be observed experimentally.

## Data availability

The data pertaining to this study are available from the corresponding author upon reasonable request.

## References

1. Bardeen, J., Cooper, L. N. & Schrieffer, J. R. Theory of superconductivity. *Phys. Rev.* **108**, 1175–1204 (1957).
2. Osheroff, D. D., Richardson, R. C. & Lee, D. M. Evidence for a new phase of solid He<sup>3</sup>. *Phys. Rev. Lett.* **28**, 885–888 (1972).

3. Klitzing, K. V., Dorda, G. & Pepper, M. New method for high-accuracy determination of the fine-structure constant based on quantized Hall resistance. *Phys. Rev. Lett.* **45**, 494–497 (1980).
4. Hensgens, T. et al. Quantum simulation of a Fermi–Hubbard model using a semiconductor quantum dot array. *Nature* **548**, 70–73 (2017).
5. Dehollain, J. P. et al. Nagaoka ferromagnetism observed in a quantum dot plaquette. *Nature* **579**, 528–533 (2020).
6. Atala, M. et al. Direct measurement of the Zak phase in topological Bloch bands. *Nat. Phys.* **9**, 795–800 (2013).
7. Meier, E. J., An, F. A. & Gadway, B. Observation of the topological soliton state in the Su–Schrieffer–Heeger model. *Nat. Commun.* **7**, 13986 (2016).
8. St-Jean, P. et al. Lasing in topological edge states of a one-dimensional lattice. *Nat. Photon.* **11**, 651–656 (2017).
9. Chaunsali, R., Kim, E., Thakkar, A., Kevrekidis, P. & Yang, J. Demonstrating an *in situ* topological band transition in cylindrical granular chains. *Phys. Rev. Lett.* **119**, 024301 (2017).
10. de Léséleuc, S. et al. Observation of a symmetry-protected topological phase of interacting bosons with Rydberg atoms. *Science* **365**, 775–780 (2019).
11. Cai, W. et al. Observation of topological magnon insulator states in a superconducting circuit. *Phys. Rev. Lett.* **123**, 080501 (2019).
12. Nielsen, E. & Bhatt, R. N. Nanoscale ferromagnetism in nonmagnetic doped semiconductors. *Phys. Rev. B* **76**, 161202 (2007).
13. Le, N. H., Fisher, A. J. & Ginossar, E. Extended Hubbard model for mesoscopic transport in donor arrays in silicon. *Phys. Rev. B* **96**, 245406 (2017).

14. Le, N. H., Fisher, A. J., Curson, N. J. & Ginossar, E. Topological phases of a dimerized Fermi–Hubbard model for semiconductor nano-lattices. *npjQuantum Inf.* **6**, 24 (2020).
15. Imada, M., Fujimori, A. & Tokura, Y. Metal–insulator transitions. *Rev. Mod. Phys.* **70**, 1039–1263 (1998).
16. Qin, M., Schäfer, T., Andergassen, S., Corboz, P. & Gull, E. The Hubbard model: a computational perspective. *Annu. Rev. Condens. Matter Phys.* **13**, 275–302 (2022).
17. Feynman, R. Simulating physics with computers. *Int. J. Theor. Phys.* **21**, 46–488 (1982).
18. Cirac, J. I. & Zoller, P. Goals and opportunities in quantum simulation. *Nat. Phys.* **8**, 264–266 (2012).
19. Argüello-Luengo, J., González-Tudela, A., Shi, T., Zoller, P. & Cirac, J. I. Analogue quantum chemistry simulation. *Nature* **574**, 215–218 (2019).
20. Altman, E. et al. Quantum simulators: architectures and opportunities. *PRX Quantum* **2**, 017003 (2021).
21. Su, W. P., Schrieffer, J. R. & Heeger, A. J. Solitons in polyacetylene. *Phys. Rev. Lett.* **42**, 1698–1701 (1979).
22. Tasaki, H. The Hubbard model—an introduction and selected rigorous results. *J. Phys Condens. Matter* **10**, 4353–4378 (1998).
23. Salfi, J. et al. Quantum simulation of the Hubbard model with dopant atoms in silicon. *Nat. Commun.* **7**, 11342 (2016).
24. Wang, X., Yang, S. & Das Sarma, S. Quantum theory of the charge-stability diagram of semiconductor double-quantum-dot systems. *Phys. Rev. B* **84**, 115301 (2011).
25. Dusko, A., Delgado, A., Saraiva, A. & Koiller, B. Adequacy of Si:P chains as Fermi–Hubbard simulators. *npjQuantum Inf.* **4**, 1 (2018).

26. Maier, T., Jarrell, M., Pruschke, T. & Keller, J. *d*-wave superconductivity in the Hubbard model. *Phys. Rev. Lett.* **85**, 1524–1527 (2000).
27. Hirsch, J. E. & Tang, S. Antiferromagnetism in the two-dimensional Hubbard model. *Phys. Rev. Lett.* **62**, 591–594 (1989).
28. Arovas, D. P., Berg, E., Kivelson, S. & Raghu, S. The Hubbard model. *Annu. Rev. Condens. Matter Phys.* **13**, 239–274 (2022).
29. Nichol, J. M. et al. High-fidelity entangling gate for double-quantum-dot spin qubits. *npj Quantum Inf.* **3**, 3 (2017).
30. Mills, A. R. et al. Shuttling a single charge across a one-dimensional array of silicon quantum dots. *Nat. Commun.* **10**, 1063 (2019).
31. Yang, C. H. et al. Operation of a silicon quantum processor unit cell above one kelvin. *Nature* **580**, 350–354 (2020).
32. MacQuarrie, E. R. et al. Progress toward a capacitively mediated CNOT between two charge qubits in Si/SiGe. *npj Quantum Inf.* **6**, 81 (2020).
33. Asbóth, J. K., Oroszlány, L. & Pályi, A. *The Su–Schrieffer–Heeger (SSH) Model* (Springer, 2016).
34. Wang, Y., Chen, C.-Y., Klimeck, G., Simmons, M. Y. & Rahman, R. All-electrical control of donor-bound electron spin qubits in silicon. Preprint at <https://arxiv.org/abs/1703.05370> (2017).
35. Watson, T. F. et al. Transport in asymmetrically coupled donor-based silicon triple quantum dots. *Nano Lett.* **14**, 1830–1835 (2014).
36. He, Y. et al. A two-qubit gate between phosphorus donor electrons in silicon. *Nature* **571**, 371–375 (2019).
37. Kranz, L. et al. Exploiting a single-crystal environment to minimize the charge noise on qubits in silicon. *Adv. Mater.* **32**, 2003361 (2020).

38. Wang, X. et al. Quantum simulation of an extended Fermi-Hubbard model using a 2D lattice of dopant-based quantum dots. Preprint at <https://arxiv.org/abs/2110.08982> (2021).
39. Kim, E. et al. Quantum electrodynamics in a topological waveguide. *Phys. Rev.* **11**, 011015 (2021).
40. Lin, S., Zhang, L., Tian, T., Duan, C.-K. & Du, J. Dynamic observation of topological soliton states in a programmable nanomechanical lattice. *Nano Lett.* **21**, 1025–1031 (2021).
41. Kruk, S. et al. Edge states and topological phase transitions in chains of dielectric nanoparticles. *Small* **13**, 1603190 (2017).
42. Rueß, F. J. et al. Realization of atomically controlled dopant devices in silicon. *Small* **3**, 563–567 (2007).
43. Broome, M. A. et al. Two-electron spin correlations in precision placed donors in silicon. *Nat. Commun.* **9**, 980 (2018).
44. Fricke, L. et al. Coherent control of a donor-molecule electron spin qubit in silicon. *Nat. Commun.* **12**, 3323 (2021).
45. Kiršanskas, G., Pedersen, J. N., Karlström, O., Leijnse, M. & Wacker, A. QmeQ 1.0: an open-source python package for calculations of transport through quantum dot devices. *Comput. Phys. Commun.* **221**, 317–342 (2017).
46. Lin, H. Q., Gubernatis, J. E., Gould, H. & Tobochnik, J. Exact diagonalization methods for quantum systems. *Comput. Phys.* **7**, 400 (1993).
47. Hogg, M. *Scalable Readout and Control of Phosphorus Donor Spin Qubits in Silicon*. PhD thesis, UNSW Sydney Australia (2021).
48. Koiller, B., Hu, X. & Das Sarma, S. Exchange in silicon-based quantum computer architecture. *Phys. Rev. Lett.* **88**, 027903 (2002).

49. Wellard, C. J. et al. Electron exchange coupling for single-donor solid-state spin qubits. *Phys. Rev. B* **68**, 195209 (2003).
50. Joecker, B. et al. Full configuration interaction simulations of exchange-coupled donors in silicon using multi-valley effective mass theory. *New J. Phys.* **23**, 073007 (2021).

## Acknowledgements

The research was supported by Silicon Quantum Computing Pty Ltd and the Australian Research Council Centre of Excellence for Quantum Computation and Communication Technology (project number CE170100012). M.Y.S. acknowledges an Australian Research Council Laureate Fellowship.

## Author information

### Author notes

1. Y. He

Present address: Shenzhen Institute for Quantum Science and Engineering, Southern University of Science and Technology, Shenzhen, China

## Authors and Affiliations

1. Centre of Excellence for Quantum Computation and Communication Technology, School of Physics, UNSW Sydney, Kensington, New South Wales, Australia

M. Kiczynski, S. K. Gorman, H. Geng, M. B. Donnelly, Y. Chung, Y. He, J. G. Keizer & M. Y. Simmons

2. Silicon Quantum Computing Pty Ltd, UNSW Sydney, Kensington, New South Wales, Australia

M. Kiczynski, S. K. Gorman, H. Geng, M. B. Donnelly, Y. Chung, J. G. Keizer & M. Y. Simmons

## Contributions

S.K.G., Y.H. and M.Y.S. conceived the project. M.K., H.G., M.B.D., J.G.K. and Y.C. fabricated the devices. M.K. and S.K.G. performed the measurements, analysed the data and performed the theoretical calculations. The manuscript was written by M.K., S.K.G., J.G.K. and M.Y.S. with input from all authors. M.Y.S supervised the project.

## Corresponding author

Correspondence to [M. Y. Simmons](#).

## Ethics declarations

## Competing interests

The authors declare no competing interests.

## Peer review

## Peer review information

*Nature* thanks Ingmar Swart and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. [Peer reviewer reports](#) are available.

## Additional information

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

## Extended data figures and tables

## Extended Data Fig. 1 Experimental measurement setup.

**a**, Schematic of the experimental set up for Device I, showing all electrical connections to the device. A STM image of the device I is shown. The gate voltages are controlled by a NIDAQ and three Yokogawas, and the source voltage by the NIDAQ. The drain current (green line) is amplified via a FEMTO low noise amplifier and acquired by the NIDAQ. **b**, Schematic of the experimental setup for Device II, showing all electrical connections to the device. A STM image of the 10 dot array is shown, taken before patterning of the control gates. The gate voltages are controlled by two NIDAQs, and the source voltage controlled by a SR830 DSP lock-in amplifier. The drain current (green line) is amplified via a FEMTO low noise amplifier and acquired by the NIDAQ.

## Extended Data Fig. 2 Energies of the many-body states in the interacting SSH model around quarter-filling.

The energies of  $m + 4$ ,  $m + 5$ , and  $m + 6$  electron states for an array of 10 sites as a function of the tunnel coupling ratio,  $v/w$ . The energies are calculated at quarter-filling (that is, the ground state is always  $m + 5$  electrons). The ground state evolves from a singly degenerate state for  $v/w > 1$  (bulk-like states) to a two-fold degenerate  $m + 5$  electron state for  $v/w < 1$  (edge-like states). The quarter-filling energy gap between the  $m + 4$  and  $m + 6$  electron ground states reduces as  $v/w < 1$  resulting in nearly degenerate states, reflecting the almost zero-energy cost associated with adding an electron to the topological phase of the array, with a small energy gap due to the inter-site Coulomb interactions.

## Extended Data Fig. 3 Splitting of the topological phase due to the presence of inter-site Coulomb interactions.

A comparison of the theoretically calculated conductance traces (blue lines) with the experimental data observed (circles) as the inter-site Coulomb interaction strength is varied from, **a**, no Coulomb interactions, to **b**,  $0.5 \times V_{i,j}$  **c**,  $V_{i,j}$  and **d**,  $2 \times V_{i,j}$ . In the case of no Coulomb interactions only a single peak is observed in the theoretical conductance trace, which does not

match the roughly 0.2 meV splitting observed in the experimental data. As the Coulomb interaction strength is increased two peaks evolve with the splitting between the peaks increasing, with the roughly 0.2 meV splitting matching the experimental data in **c**.

## Supplementary information

### Supplementary Information

This Supplementary Information file includes Supplementary Tables 1–6, Figs. 1–5 and References.

### Peer Review File

## Rights and permissions

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

### Reprints and Permissions

## About this article

### Cite this article

Kiczynski, M., Gorman, S.K., Geng, H. *et al.* Engineering topological states in atom-based semiconductor quantum dots. *Nature* **606**, 694–699 (2022). <https://doi.org/10.1038/s41586-022-04706-0>

- Received: 05 November 2021
- Accepted: 01 March 2022
- Published: 22 June 2022
- Issue Date: 23 June 2022
- DOI: <https://doi.org/10.1038/s41586-022-04706-0>

## Share this article

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

[Get shareable link](#)

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

[Copy to clipboard](#)

Provided by the Springer Nature SharedIt content-sharing initiative

---

This article was downloaded by **calibre** from <https://www.nature.com/articles/s41586-022-04706-0>

- Article
- Open Access
- [Published: 22 June 2022](#)

# Organic bipolar transistors

- [Shu-Jen Wang](#),
- [Michael Sawatzki](#),
- [Ghader Darbandy](#),
- [Felix Talnack](#),
- [Jörn Vahland](#),
- [Marc Malfois](#),
- [Alexander Kloes](#),
- [Stefan Mannsfeld](#),
- [Hans Kleemann](#) &
- [Karl Leo](#)

*Nature* volume **606**, pages 700–705 (2022)

- 599 Accesses
- 152 Altmetric
- [Metrics details](#)

## Abstract

Devices made using thin-film semiconductors have attracted much interest recently owing to new application possibilities. Among materials systems suitable for thin-film electronics, organic semiconductors are of particular interest; their low cost, biocompatible carbon-based materials and deposition by simple techniques such as evaporation or printing enable organic semiconductor devices to be used for ubiquitous electronics, such as those used on or in the human body or on clothing and packages<sup>1,2,3</sup>. The potential

of organic electronics can be leveraged only if the performance of organic transistors is improved markedly. Here we present organic bipolar transistors with outstanding device performance: a previously undescribed vertical architecture and highly crystalline organic rubrene thin films yield devices with high differential amplification (more than 100) and superior high-frequency performance over conventional devices. These bipolar transistors also give insight into the minority carrier diffusion length—a key parameter in organic semiconductors. Our results open the door to new device concepts of high-performance organic electronics with ever faster switching speeds.

## Main

Organic field-effect transistors (FET) were first reported in 1986 and have shown impressive improvements in the past two decades<sup>4,5,6,7,8,9,10,11</sup>. Nevertheless, they are still restricted to the low-to-medium megahertz range, which does not allow broad application<sup>12,13,14</sup>. The substantially lower charge carrier mobility in organic semiconductors (OSCs) compared with their inorganic counterparts is a limitation to the performance of organic transistors. Reducing the length of transistor channels is an effective strategy for improving the operational speed of the device, as shown both in FET<sup>13,14</sup> and other device concepts such as organic permeable-base transistors<sup>11,15</sup>. However, other factors, such as contact resistance and overlap capacitances, often limit further improvement of operational frequencies<sup>16,17</sup>.

A device that offers both low capacitance and contact resistance is the bipolar junction transistor. Although they have disadvantages with regard to miniaturization and process integration, bipolar transistors possess substantially higher operational speeds than comparable field-effect devices<sup>18</sup>. However, organic bipolar junction transistors (OBJTs) have not yet been realized, mainly because they rely on minority carrier diffusion through a thin and precisely doped base layer. Most studies have addressed exciton diffusion, which dominates owing to the weak dielectric screening in organic compounds<sup>19,20</sup>. Majority carrier diffusion length in fullerenes has been estimated to be on the centimetre scale, raising interesting questions about carrier diffusion physics in OSCs<sup>21,22</sup>. Charge carrier minority diffusion lengths have remained unexplored in OSC materials until now. In

comparison to exciton diffusion, they can be expected to be in the nanometre range, at least for typical disordered organic films<sup>23,24,25</sup>.

Here, we realize an OBJT based on crystalline films of n- and p-type doped rubrene. In contrast to common furnace-grown single crystals, these films are made directly on the surface of a substrate and are thus compatible with mass production. We have demonstrated previously the excellent device potential of such highly ordered films by showing record-high vertical charge carrier mobilities that enabled ultrafast diode devices to operate in the gigahertz range<sup>26</sup>. Here we demonstrate that OBJTs based on crystalline rubrene thin films provide a promising route towards gigahertz organic electronics. Numerical simulations clarify the principles of transistor operation and present routes towards further optimization. A careful analysis of the device operation enables the direct measurement of minority carrier diffusion length in any OSC.

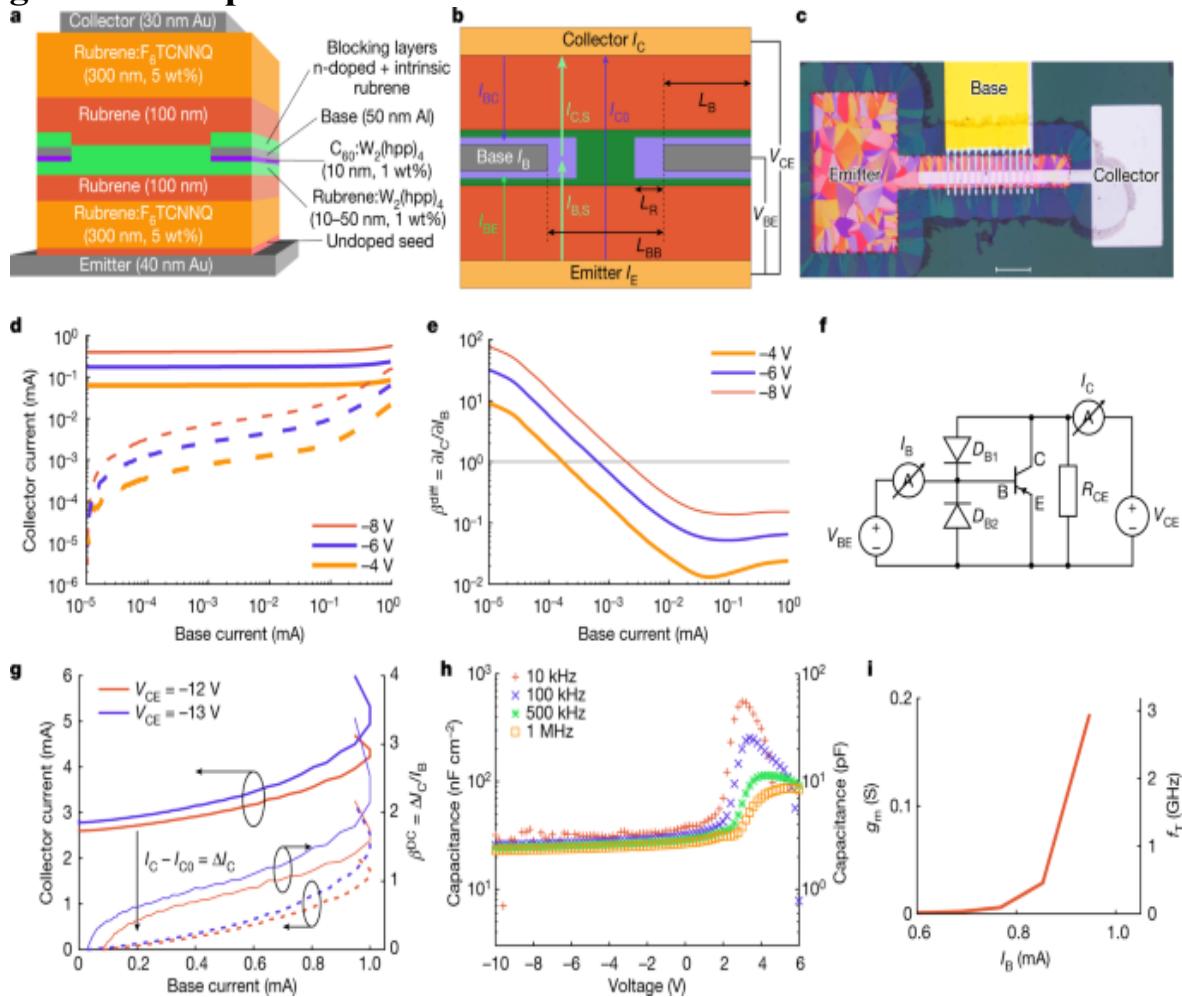
A key challenge in realizing an organic bipolar transistor is to find a suitable material and a device configuration that (1) allow both n- and p-type doping; (2) have sufficiently high (more than  $1 \text{ cm}^2 \text{ V}^{-1} \text{ s}^{-1}$ ) mobility allowing for balanced hole and electron transport, giving hope that the, so far unknown, minority carrier diffusion lengths are high enough to allow the carriers to travel through the base layers; and (3) allow a sufficiently thin base held at a defined potential to allow emitter–collector current control. We made use of the highly crystalline rubrene thin-film crystals with n- and p-type doping for the construction of this OBJT and analyse its operation experimentally and theoretically (for details of materials development and characterization, see [Methods](#)).

## Development of OBJTs

Using these highly crystalline doped films, we produced an OBJT. The device geometry is shown in Fig. [1a–c](#), featuring a vertical stacking of a rectangular emitter electrode at the bottom, a finger-like structured base electrode in the middle and a rectangular collector (top) electrode. The distancing between adjacent fingers of the base electrode and the width of each base finger itself are crucial, as discussed below. The final device is of pnp type, with an n-doped base, as we expect the p-type minority diffusion

length to be higher owing to higher mobility. As is common for organic diode-like devices<sup>26</sup>, intrinsic films are added in between p- and n-doped films to improve reverse leakage behaviour, ending up with a pinip structure. Emitter and collector electrodes are made from gold to facilitate efficient hole injection, whereas the base electrode consists of aluminium for better electron injection. A thin film of n-doped C<sub>60</sub> is added on the emitter side of the base electrode to further facilitate electron injection. Additional layers of intrinsic and weakly doped material can be added on top of the base electrode to minimize base–collector leakage.

**Fig. 1: OBJT operation.**



**a**, Vertical stack configuration of the OBJT. **b**, Definition of active and parasitic currents and lateral geometric parameters in the OBJT. **c**, The OBJT device under a polarized microscope. Scale bar, 100  $\mu\text{m}$ . **d**, Transfer characteristics of the OBJT device with blocking layers deposited on top

of the base electrode for different  $V_{CE}$ : solid lines give the absolute collector current  $I_C$ , dashed lines give added current  $\Delta I_C = I_C - I_{C0}$ . **e**, The corresponding differential amplification for the device in **d**. **f**, Definition of the biasing and measurement setup for all OBJT curves and representation of the equivalent circuit of the OBJT containing active and parasitic components analogous to the currents defined in **b**:  $D_{B1}$ , direct base–collector diode with  $I_{BC}$  leakage current;  $D_{B2}$ , direct base–emitter diode with  $I_{BE}$  leakage current;  $R_{CE}$ , direct emitter–collector overlap with  $I_{C0}$  output off-current. **g**, Transfer characteristics of the OBJT device without blocking layers deposited on top of the base electrode at different  $V_{CE}$ : solid thick lines, absolute collector current  $I_C$ ; dashed lines, added current  $\Delta I_C = I_C - I_{C0}$ ; solid thin lines, absolute direct current amplification. **h**, Absolute and area-normalized capacitance of an individual rubrene-based pin (input) diode at different biasing conditions and varying measurement frequencies. The active area is  $100 \times 100 \mu\text{m}^2$ . **i**, Transition frequency estimation from transconductance.

It might seem self-evident to use the triclinic crystal phase of rubrene for bipolar junction transistors owing to their higher vertical charge carrier mobility, facilitating a more efficient vertical diffusion through the base layer. However, in addition to vertical transport, the n-doped rubrene layer of the base should be an area of equipotential with the metallic base electrode, which requires a high lateral conductivity. The distance between adjacent metallic base electrodes is the defining geometric parameter for this device concept and is in the range of micrometres. Therefore, orthorhombic crystals are used here successfully for OBJT because of their isotropic charge transport properties—transistor operation using triclinic crystals was not observed.

We first look at a device based on orthorhombic spherulite crystals with more blocking layers deposited on top of the base electrode (Fig. [1d,e](#)). The base–emitter diode, the base–collector diode and the emitter–collector pinip structure are first investigated separately to check functionality at component level (Extended Data Fig. [1a](#)). The input and output components function individually as diodes, with distinguishable forward and reverse behaviour. The base–collector diode possesses a substantially lower forward

current than the base–emitter input diode owing to the extra blocking layers on top of the base electrode. However, the reverse current and forward leakage on both sides of the base are almost identical. This is a sign that the leakage current is governed by lateral leakage paths rather than the current going through the pin diodes. As expected, the direct current from emitter to collector is fully symmetric (see impedance measurements in Extended Data Fig. 2). However, this current is substantially higher than the current through the diodes themselves. The high emitter–collector current can be explained partly by the simple electrode design, which creates a large area of parasitic overlap between emitter and collector. It is possible to reduce the emitter–collector current by structuring the electrode. A discussion about the optimal geometric configuration based on simulations is given in the next section. Our main focus here is on the base region to enable the operation of the OBJT.

Figure 1d shows the transfer curve of the full OBJT (the electrode gap in base electrode is  $12\text{ }\mu\text{m}$ ), that is, the emitter (output) current over the base (input) current at different emitter–collector voltages. It is obvious that the absolute value of the emitter current is large and barely changes throughout the measurement. Only at high base currents is a slight increase noticeable. This is caused by the emitter–collector leakage current discussed above. This leakage current can be seen as a constant shunt  $R_{CE}$  in parallel with the output of the device (an equivalent circuit is presented in Fig. 1f). Thus, the real output of the transistor reflects the change in collector current (also shown in Fig. 1d) controlled by the base current. A steady increase in output current over input current is visible, with a steep increase at low and high base currents and a substantially shallower slope in the medium current regime. The general behaviour is similar for all applied emitter–collector voltages, albeit shifted by an absolute current. Focussing on the largest  $V_{CE}$  of  $-8\text{ V}$ , the added collector current surpasses the input base current only until a base current of  $15\text{ }\mu\text{A}$ .

Figure 1e shows the differential signal amplification  $\partial I_C / \partial I_B$ . It is as large as 100 at a low base current, clearly proving transistor action, and then decreases steadily with increasing base current. The loss of differential amplification occurs at  $I_B = 2\text{ }\mu\text{A}$ . This decrease in differential amplification can be understood from the geometry of the device: an illustration of the

current paths is given in Fig. 1b. In addition to the already mentioned current path through  $R_{CE}$ , leading to a large  $I_{C0}$ , the top and bottom diodes of the transistor can be split into two parts. First, a large part of each diode is defined by the area of direct overlap of the base and the collector or emitter electrode. This region contributes only to the leakage current and does not contribute to transistor operation. The leakage current through the base–collector diode  $D_{B1}$  is denoted as  $I_{BC}$  and the leakage current through the base–emitter diode  $D_{B2}$  as  $I_{BE}$ . Second, a smaller part is defined by the area around the base electrode fingers in which the base potential is present. This distance is given by the base reach  $L_R$ . The corresponding area is marked in Fig. 1b. Only the second part (current  $I_{B,S}$ ) can contribute to the modulation of the collector current in the form of  $I_{C,S}$ . The equivalent circuit of this configuration is shown in Fig. 1f. The amplification of the transistor component starts to saturate at higher base currents owing to the exponential increase in input current through the parasitic parts of the input diode such that differential amplification cannot be maintained at higher base current. The measured differential amplification is therefore not an intrinsic property of the transistor but a property of the device functioning as a circuit.

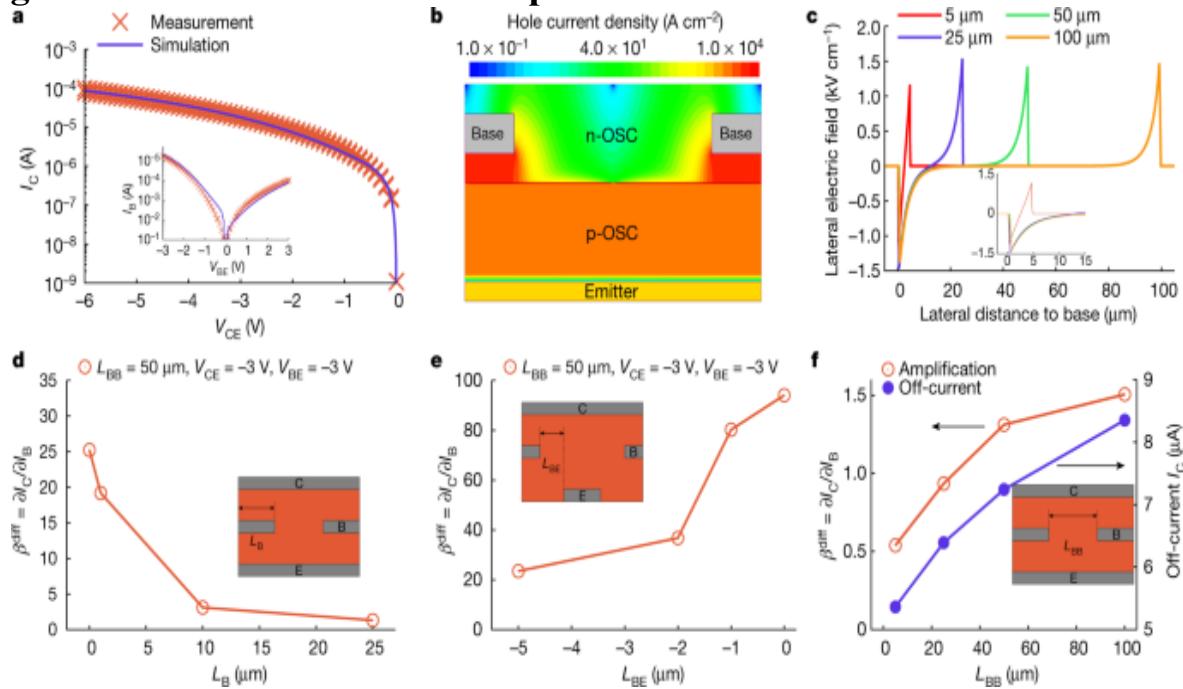
The blocking layers deposited on top of the base electrode aimed at suppressing leakage current from the parasitic diode  $D_{B1}$  can, however, also block part of the channel next to the base electrode fingers that would contribute to the transistor operation due to the geometry of thermal evaporation through shadow masks. A balance must be found between the configuration of blocking films and electrode geometry. We also investigated a device based on orthorhombic platelet crystals without the use of blocking layers on top of the base electrode, the current–voltage ( $IV$ ) characteristics of the device are shown in Fig. 1g. The transistor operation of the device at low currents is reduced, because the changed biasing of the base and the resulting change in parasitic current of the diode  $D_{B1}$  overcompensate for any diffusion-based amplification. However, at high base currents, the output collector current is increased substantially, and the transistor clearly shows large-signal amplification, although only moderate values. We would like to note that the unstable behaviour at high base current is probably caused by the high current density in the device, which is close to the onset of self-heating effect. Therefore, both differential and

absolute current amplification can be observed in our OBJT devices based on doped rubrene crystals.

## TCAD simulations of OBJTs

Technology computer-aided design (TCAD) simulations are performed to obtain a better understanding of the charge transport in the OBJT device and design rules for optimization of the device geometry. The simulations are based on the device stack that showed large-signal amplification as shown in Fig. 1g. The fabricated devices and experimental data are taken as a reference to calibrate the TCAD simulator. The  $IV$  characteristics of the individual components (base–emitter diode and emitter–collector structure) show good agreement between the calibrated simulation results and measured data as shown in Fig. 2a, thus confirming the viability of the device operation. On the basis of the calibrated TCAD, electrostatic potential, electric field, carrier density and the current, distributions can be simulated and extracted for different bias conditions and geometries. As an example, Fig. 2b shows the current density distribution in the OBJT. The lateral electric field distribution between two adjacent base fingers is shown in Fig. 2c.

**Fig. 2: TCAD simulation of the operation of the OBJT device.**



**a**, Congruence of simulation and measurement on the basis of the experimental data from Fig. 1g. The simulation is tuned to reproduce *IV* characteristics of the emitter–collector and emitter–base (inset) individually. **b**, The geometry and current density distribution for an exemplary configuration of the OJBT as given by TCAD simulations. **c**, Field strength of the internal electric field in a lateral direction for different distances between adjacent base electrodes at  $V_{BE} = V_{CE} = -3$  V. The inset shows a close-up view of the panel for clarity. **d**, Simulated maximum differential amplification with different widths of the base electrode  $L_B$ . **e**, Simulated maximum differential amplification with hypothetical lateral offset between the end of the base electrode and the start of the emitter electrode  $L_{BE}$ . **f**, Simulated maximum differential amplification with different distances between adjacent base electrodes  $L_{BB}$  (all other parameters were kept constant in each set of simulations). The insets show the geometry of  $L_B$ ,  $L_{BE}$  and  $L_{BB}$  in the OJBT device. The parameters used are summarized in Supplementary Table 1.

The simulation provides an insight into a key parameter of the transistor: the finger design of the base electrode. The length required for the field to drop from its maximum value to almost zero can be interpreted as the base reach  $L_R$ . For an electrode distance of more than 25  $\mu\text{m}$ , the lateral field is close to zero for an important part of the device, causing a large initial off-current that is not controllable by the base current. On the basis of the simulation, a base-to-base distance of 5  $\mu\text{m}$  to 10  $\mu\text{m}$  seems to be optimal.

Figure 2d shows the impact of the size and arrangement of the base electrodes on amplification. When the width of the base electrode is reduced from 25  $\mu\text{m}$  to, theoretically, 0  $\mu\text{m}$  (this is equivalent to no direct overlap between base, emitter and collector), the maximum amplification increases substantially, as the part of the base current that does not contribute to amplification decreases, whereas the controllable collector current remains the same. However, a base overlap of 0  $\mu\text{m}$  is impossible to achieve for technological reasons. By contrast, a negative overlap in the sense of a spacer/gap between the end of the base and the beginning of the emitter can be achieved. Figure 2e shows the resulting amplification for such configuration. The amplification is reduced as expected as the important

edge area near the base electrode is now substantially less involved in the transport. However, the reduction is comparably modest for a gap length of 1  $\mu\text{m}$ . This degree of alignment precision would be achievable with advanced stencil lithography techniques.

Finally, the distance between adjacent base electrodes is varied, as shown in Fig. 2f. Surprisingly, the amplification increases slightly for increased distances between adjacent base electrodes, although a saturation is seen above 50  $\mu\text{m}$ . This is because, although the lateral field is close to zero far from the base (Fig. 2c), a small amount is still contributing to the output current. However, the off-current is also increased when the distance between bases is increased because the emitter–collector overlap increases simultaneously (Fig. 2f). Therefore, there is a trade-off between the current amplification and the off-current when designing the base electrode.

Overall, the simulations confirm the operation of the OBJs with differential as well as large-signal amplification. Furthermore, they give clear design guidelines how to further improve the devices.

## Operation speed of OBJs

With a total device thickness of approximately 1  $\mu\text{m}$  and a high vertical mobility of approximately  $3 \text{ cm}^2 \text{ V}^{-1} \text{ s}^{-1}$ , OBJs seem well suited for high-frequency operation. The most important dynamic performance parameter for any kind of transistor is the unity-gain cut-off frequency. A direct measurement of this quantity requires sufficient large-signal amplification and stability of operation. Unfortunately, in our OBJs, we obtain large-signal amplification only at the highest applied bias, which results in unstable behaviour (Fig. 1g). Still, we reasonably estimate the maximum speed of operation by evaluating the resistor–capacitor time of the system. Similar to the calculations done for FET, it is possible to estimate the maximum speed of operation in the form of the transition frequency from static properties using:

$$\text{\$}\{f\}_{\text{\rm T}}=\frac{\text{\rm g}_m}{2\pi C} \quad (1)$$

where  $g_m$  and  $C$  denote the transconductance of the transistor and the capacitance, respectively. The transconductance describes the change in output current with input voltage. In case of the OBJT, it can be written as:

$$\begin{aligned} \text{\$\$}\{g\}_{\{\{\rm m\}\}} &= \frac{\partial I}{\partial V_{\rm BE}} \\ &= \beta \frac{\partial I}{\partial V_{\rm BE}} \end{aligned} \quad (2)$$

Because the output current ( $I_C$ ) is linked to the input current ( $I_B$ ) through the amplification ( $\beta$ ), the transconductance is defined by the differential conductance of the input diode. Similarly, the defining capacitance is given by the input diode, assuming diffusion through the base is sufficiently fast, the transition frequency is seemingly limited only by the properties of the input diode. On the basis of the results obtained from the simulations, one goal is to reduce the direct base current as much as possible, which would reduce conductance of the input diode. However, the amplification of the device would increase accordingly, leaving the  $g_m$  constant. A direct transition frequency measurement is challenging for OBJTs owing to the parasitic diodes that influence the phase of small signal measurements. Nevertheless, the high degree of agreement between the direct transition frequency measurements and the transconductance/capacitance estimations in the literature allow us to estimate the frequency response of our OBJTs<sup>13,14,27</sup>. For the device shown in Fig. 1g, the resulting transconductance is as high as 0.1 S (Fig. 1i), in the range in which devices show amplification, whereas the capacitance is around 10 pF (Fig. 1h). This results in a transition frequency of 1.6 GHz, which is similar to the speed of operation found for the single, rubrene-based diodes<sup>26</sup> and hence provides a significant step (10–40×) above the current state of the art of organic transistors<sup>12,14</sup>. Two reasons for the superiority of the OBJT are (1) highly crystalline films that feature improved mobilities compared with most OSCs and (2) the ultralow capacitance of devices associated with the vertical bipolar junction transistor design. In addition, limitations to contact resistance are less prominent here, because all the metal–OSC interfaces are doped by default and do not limit injection, proven by the space-charge-limited current analysis-like behaviour in rubrene pip devices.

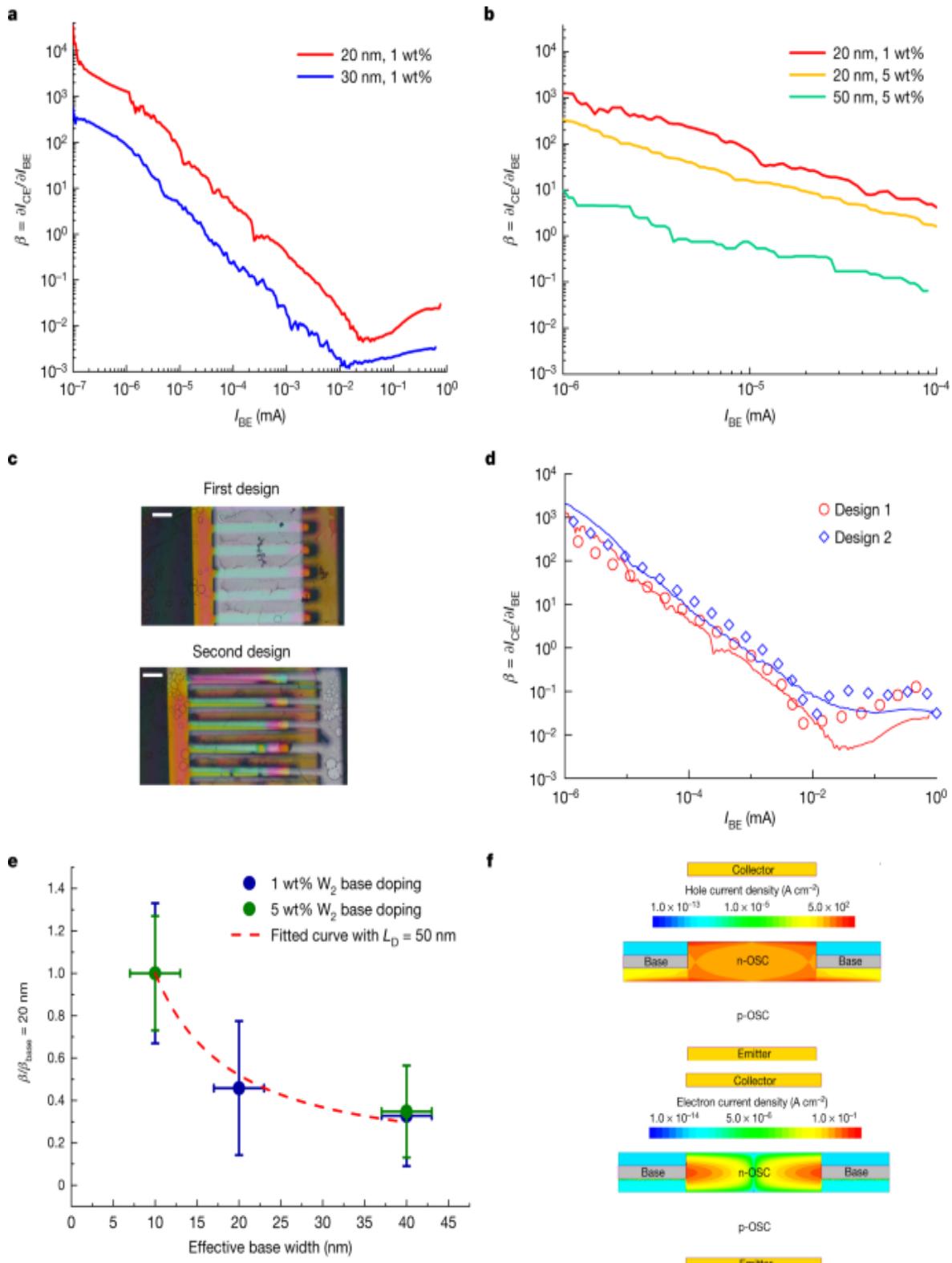
## Minority carrier diffusion length

The working principle of the OBJT is based on the diffusion of minority carriers (holes) through the base (n-doped film). In an ideal device, the diffusion length could be calculated directly from the doping concentrations, the width of the base layer and the resulting amplification. However, as discussed, the amplification measured here does not represent the intrinsic amplification of the transistor itself but of the device as a circuit.

Nevertheless, the observation of amplification proves the diffusion of minority carriers through the base, with a minority diffusion length of at least 20 nm for devices with 1 wt% of base doping. In addition, we conducted experiments for which we varied the properties of the doping and structure of the base. Consistent with the inorganic bipolar junction transistor theory, both an increase in base doping from 1 wt% to 5 wt% and an increase in base layer thickness substantially reduce current amplification. The strong dependency of OBJTs on the base thickness and doping concentration is associated with the minority carrier diffusion operation, which is in stark contrast to organic permeable-base transistor operation based on most carrier transport. It is possible to estimate the diffusion length from devices with different base thickness when all remaining parameters remain identical. Figure 3 shows the OJBT operation based on a new set of devices with improved electrode geometry, reducing the area of electrode overlap that does not contribute to the transistor operation. The reduction in the parasitic electrode overlap area improves transistor performance, which is in line with the TCAD simulations (Fig. 3d). On the basis of these measurements, the diffusion length for holes through the n-doped rubrene is estimated, by fitting the classical bipolar transition relation  $\beta \propto \frac{W}{L_D}$  together with the calibrated TCAD simulation, to be roughly 50 nm, showing excellent agreement with experimental results and minority-carrier-dominated device operation by using an input diffusion length of 50 nm (Fig. 3d-f and Extended Data Figs. 1 and 3). Exciton diffusion lengths in the micrometre range found in photoexcitation experiments on single crystals of rubrene<sup>28</sup> indicate fundamentally different mechanisms governing the transport and relaxation of minority holes. Considering the high structural order of rubrene crystals after doping, the recombination processes are probably caused by the slight widening of the

density states. Our OBJT device provides a tool to obtain direct access to the physical properties of minority carrier diffusion in similarly high mobility OSC systems, opening the possibility to investigate fundamental questions about mechanisms of minority recombination in OSCs.

**Fig. 3: Thickness and doping concentration of OBJTs.**



**a,b**, Differential amplification of OBJTs with different base layer thicknesses (**a**) and tetrakis(hexa-hydropyrimidinopyrimidine)ditungsten(II)

(W<sub>2</sub>(hpp)<sub>4</sub>) doping concentrations (**b**). **c,d**, Optical microscope images of different OBJT electrode designs (**c**) and corresponding device differential amplification curves (**d**). The solid lines denote experimental results and the hollow symbols denote TCAD simulation results. Scale bars, 100 μm. The device has a base thickness of 20 nm with 1 wt% W<sub>2</sub>(hpp)<sub>4</sub> doping concentration. **e**, Normalized differential amplification as a function of effective base width and doping. The effective base width is the base thickness minus the space charge length (2LSCL is approximately 10 nm determined from TCAD simulation and electrical characterization). The differential amplification was taken at a base current of 10<sup>-5</sup> mA, for which there is a good agreement between the TCAD and experiment results. The error bars denote the standard error of the mean by averaging over five devices prepared in a single run. The red curve is a coth fit with minority diffusion length of 50 nm. **f**, TCAD-simulated hole current density as a minority carrier (top) and electron current density (bottom) in the n-doped base layer.

In summary, we demonstrate a functional OBJT, delivering a missing piece of the puzzle on the organic transistor roadmap. Our OJTs, based on highly crystalline rubrene thin-film crystals, not only provide a promising route towards ultrahigh-frequency organic transistors, but also allow the study of important fundamental physical parameters such as the minority carrier diffusion length, estimated to be around 50 nm for a doping concentration of 5 wt% for rubrene crystals. We believe that our results pave the way for next-generation high-performance organic electronic devices and provide a tool for understanding carrier diffusion physics in high mobility OSCs.

## Methods

### Details of rubrene thin-film crystal development

#### Growth procedures

The general process for growing thin-film crystals of rubrene was described in refs. [29,30](#). In Extended Data Fig. [4a,b](#), we show the fabrication process for rubrene thin-film crystals and types of rubrene thin-film crystal phase upon

doping, respectively. A thin layer of amorphous rubrene is deposited on a substrate by vacuum deposition and then annealed in a nitrogen atmosphere to initiate crystal growth. Different crystal phases are possible depending on surface properties and heating temperature. The three most common types of crystal are triclinic spherulites, orthorhombic spherulites and orthorhombic platelets (Extended Data Fig. 4b). Triclinic crystals start to form from approximately 120 °C and are the most robust and reproducible of the common crystal phases. Previously, we have shown the improved properties of triclinic films in gigahertz diodes<sup>26</sup>. Although the vertical mobility in these triclinic films is high, lateral transport is inefficient because of the strongly branched nature of these films. Orthorhombic crystals are the main focus in most publications owing to the isotropic charge transport properties originating from its herringbone molecular packing with ideal wavefunction overlap<sup>31,32,33</sup>. The spherulitic configuration of the orthorhombic packing grows at higher temperatures above 170 °C without strong branching, and can be identified easily under a polarized microscope by straight rays fanning out from the individual centre of each crystallite. Orthorhombic platelets are the most difficult phase to be created consistently. Heating at 150 °C to 170 °C commonly results in a few single crystals or clusters of crystals distributed over the surface. A previous study showed that a uniform and surface-covered distribution of platelet crystals can be achieved by the introduction of a sublayer with appropriate glass transition temperature<sup>34</sup>. Here we use 5 nm of 4,4'-cyclohexylidenebis[N,N-bis(4-methylphenyl)benzenamine] (TAPC), resulting in successful crystal growth on glass and silicon substrates as well as structured metal and indium-tin-oxide electrodes<sup>35</sup>.

## Epitaxy and doping

To make an OGBT, we need to control the total thickness of the crystal and the sequence of doped films precisely to realize complex device stacks. We introduced doping using coevaporation into initial seed and epitaxially grown layers. The maximum concentration of dopant that allows reproducible crystallization of the seed is below 2 wt% for both the p-type and n-type dopants studied here. Films added using epitaxy can be doped at substantially higher concentrations without any great changes in morphology visible by polarized microscopy. However, a change in surface properties

can be seen in atomic force microscopy (AFM) measurements. The plateaus intermixed with line and screw dislocations that are described in ref. [36](#) are visible for the undoped crystals but gradually change into a more granular surface with fewer distinct features when doping is introduced (Extended Data Figs. [5](#) and [6](#)).

## Structural analysis

GIWAXS measurements of thin films of seed and bulk material show a change in the molecular packing of the rubrene crystals upon doping. Two-dimensional (2D) plots of the scattering image (Extended Data Fig. [7a](#)) prove the high degree of crystallinity, especially of the orthorhombic platelet form. The widths of the corresponding scattering peaks indicate the degree of disorder along the corresponding axis. Here the in-plane signal (*xy*) corresponds to the *a*- and *b*-crystal axis of the rubrene unit cell, which is important for lateral transport. The out-of-plane axis is defined by the *c* axis, relevant for the vertical transport. Extended Data Fig. [7b](#) shows the change in peak width in both directions depending on the doping concentration and type. Peaks are substantially broader in the out-of-plane direction, which can, however, be attributed partly to the way the data are analysed (see the '[GIWAXS analysis](#)' section and Extended Data Fig. [8](#)). The relative change is, however, more important than the absolute values. The in-plane data behave as expected in that a higher doping concentration results in a broadening of peaks, indicating a reduction in molecular order. Introduction of the n-dopant tetrakis(hexahydropyrimidinopyrimidine)ditungsten(II) ( $\text{W}_2(\text{hpp})_4$ ) has a stronger impact than the p-dopant 1,3,4,5,7,8-hexafluorotetracyanonaphthoquinodimethane ( $\text{F}_6\text{-TCNNQ}$ ) when both films are doped to the same weight concentration. The introduction of dopant to the bulk part of the film generally increases the disorder in the film for both platelet and spherulitic samples (Extended Data Fig. [7b](#)). The out-of-plane axis behaves differently. Here doping of the seed shows a strong change in peak width, suggesting that integration of the dopant into the structure during seed crystallization is influencing mainly the *c* direction. Integration of dopant into the bulk films gradually increases the peak width, similar to the in-plane behaviour. However, the relatively stronger impact of the n-dopant compared with that of the p-dopant is even more pronounced. It can be concluded that the introduction of dopant molecules changes the

molecular structure of the rubrene films, but only to a limited degree. Higher doping concentrations create stronger disturbance, whereas  $\text{W}_2(\text{hpp})_4$  shows a stronger impact than F6-TCNNQ, presumably because of the size and steric properties of the three molecules.

## Charge transport

Lateral electrical transport has been studied extensively in undoped films of all three crystal phases<sup>26,29,34,35,37,38</sup>. Lateral mobilities are in the range of  $10^{-2} \text{ cm}^2 \text{ V}^{-1} \text{ s}^{-1}$  for triclinic films<sup>25</sup> and  $1\text{--}4 \text{ cm}^2 \text{ V}^{-1} \text{ s}^{-1}$  for orthorhombic films<sup>34,35,37,38</sup>. Lateral charge carrier mobility in platelets is usually slightly better than in spherulitic crystals, depending on the orientation of the crystal towards the electrode. However, in vertical organic devices, including the bipolar junction transistors investigated here, lateral and vertical transport occur simultaneously. Previously, we presented data on the vertical and lateral transport of undoped and doped films of the triclinic crystal phase<sup>25</sup>. Despite their superior transport properties in the vertical direction, these films are not suitable for OBJT devices owing to their mediocre lateral transport properties. Therefore, we will focus mainly on the vertical charge transport properties in orthorhombic crystals that are relevant to our OBJT devices.

Extended Data Fig. 4c shows  $IV$  curves of crystalline thin films of all three crystal phases for 400 nm undoped material sandwiched between gold electrodes. In contrast to the lateral measurements, vertical conduction is largest for triclinic films, whereas both orthorhombic crystal types behave similarly. This finding is expected because the stacks perpendicular to the surface are denser in the triclinic polymorph and identical for both orthorhombic crystal types. The differences between platelet and spherulitic films can be explained by the impact of injection owing to the low mobility and deep ionization potential of the TAPC sublayer used for the platelets<sup>39</sup>.

To further analyse the transport, we performed a space-charge-limited current analysis (SCLC) for the spherulite crystals based on sets of films with 400 nm and 600 nm of intrinsic crystal ( $L$ ) sandwiched between 40 nm of injection layers doped with 5 wt% of the p-dopant F6-TCNNQ and gold electrodes (Extended Data Fig. 4e). At high voltages (more than 1 V), a clear

quadratic dependence is visible, indicating the SCLC behaviour of holes. The estimated vertical mobility for spherulite crystals is around  $3 \text{ cm}^2 \text{ V}^{-1} \text{ s}^{-1}$  (see Extended Data Fig. 9a for detailed SCLC analysis), which is lower than that of the triclinic crystal phase (approximately  $10 \text{ cm}^2 \text{ V}^{-1} \text{ s}^{-1}$ )<sup>26</sup>. The difference between vertical and lateral mobility in orthorhombic crystals is close to isotropic, which is beneficial for applications in which charge transport occurs in both the lateral and vertical directions. As an illustration, Extended Data Fig. 4d (spherulites) and Extended Data Fig. 9b (platelets) show the impact of doping with F<sub>6</sub>-TCNNQ on the vertical current conduction. Even small amounts of doping increase the vertical conduction by orders of magnitude. The increased conduction at small voltages (less than 0.1 V) indicates that a significant part of this increase in conduction could be attributed to the reduction in injection resistance. A further increase in doping concentration causes a matching increase in current; however, the efficiency of the doping process decreases with higher doping concentration as expected from highly crystalline systems<sup>40</sup>. Electron doping of rubrene with the n-dopant W<sub>2</sub>(hpp)<sub>4</sub> works analogously, albeit with a lower doping efficiency and lower charge carrier mobility<sup>26,41</sup>.

## Sample preparation

Devices are fabricated on glass wafers with a size of  $25 \times 25 \text{ mm}^2$ . Substrates are cleaned in acetone, ethanol, isopropanol and deionized water. Each substrate is treated in piranha solution for 15 min to generate a clean and hydrophilic surface before being rinsed in deionized water and dried with nitrogen. Rubrene is provided by TCI, and F<sub>6</sub>-TCNNQ and W<sub>2</sub>(hpp)<sub>4</sub> are provided by Novaled. Layers are deposited using thermal evaporation under vacuum with a base pressure of  $1 \times 10^{-8} \text{ mbar}$ . The evaporation rate of the seed has no influence on the remainder of the process. After deposition of the bottom metal electrode (30–40 nm), the sublayer of TAPC (5 nm) and the first amorphous layer of rubrene (30–40 nm), samples are transferred to a nitrogen glovebox, without exposure to air. Heat treatment takes place on a preheated hotplate at 160–180 °C, for 1–3 min. If needed, more layers are added using coevaporation of rubrene and dopant with the same vacuum deposition at rates between  $0.5 \text{ \AA s}^{-1}$  and  $3 \text{ \AA s}^{-1}$ , depending on the doping concentration. Electrodes and semiconductor are structured using shadow

masks. Active areas for conductivity and SCLC measurements range from  $50 \times 50 \mu\text{m}^2$  to  $150 \times 150 \mu\text{m}^2$ . Devices used for conductivity measurements have a total thickness of 400 nm. The initial seed is undoped. No further doping other than the given bulk doping is introduced at the electrodes. SCLC was analysed using two sets of devices with 400 nm and 600 nm total thickness  $L$  each and active areas between  $50 \times 50 \mu\text{m}^2$  and  $150 \times 150 \mu\text{m}^2$ . The stack consists of 20 nm of undoped seed and the corresponding thickness of undoped bulk layer sandwiched between 40 nm of doped film (5 wt% for injection) and 30 nm of gold. The mobility value is extracted using the  $1/L^3$  dependence of the fits gained from the fits of the  $V^2$ -dependent SCLC current. For OBJT devices, silicon-based stencil masks are used to structure the metal electrodes, the emitter and collector electrodes consist of simple overlapping rectangles, whereas the base electrodes consist of a comb-like structure with rectangular fingers. The widths of the emitter and collector electrodes are 100  $\mu\text{m}$  and 60  $\mu\text{m}$ , respectively. The width of the base fingers is 12  $\mu\text{m}$ , with spacing between them kept to 12  $\mu\text{m}$ . The number of fingers in each of the comb-like structures of the base electrode is adjusted to the width and spacing of the fingers to approximately cover the overlap area between the emitter and collector electrodes. The devices used for the tests shown in Fig. 3 have a finger-like electrode that is around 15  $\mu\text{m}$  each for both the emitter and base. The collector electrode is either a standard rectangular stripe or finger-like electrode (details provided in Extended Data Fig. 10).

## Measurements

We performed electrical direct current measurements using Keithley 236, Keithley 2400 and Keithley 2600 source measure units, in which capacitance measurements were done with an HP 4284A in a nitrogen atmosphere. The electrical measurements were taken using the measurement software SweepMe! (sweep-me.net). Micrographs were taken with a Nikon Eclipse LC100 PL/DS polarization microscope. We performed AFM measurements with an AIST-NT Combiscope1000 and GIWAXS measurements at the B111 NCD-Sweet beamline at the ALBA synchrotron in Barcelona, Spain. The thin films were illuminated under a grazing angle of 0.12 with a beam energy of 12.95 keV and a beam size of  $70 \times 150 \text{ m}^2$  (vertical  $\times$  horizontal). The diffraction pattern was recorded with an LX255-

HS area detector from Rayonix, which was placed approximately 14 cm behind the samples. Chromium oxide ( $\text{Cr}_2\text{O}_3$ ) was used to calibrate the sample–detector distance and the beam position on the detector. The data were analysed with the WxDiff software (S.M.).

## TCAD simulation

Synopsys TCAD was used with advanced physical models and the device simulation tools (structure editor, sdevice, svisual and inspect) to simulate the electrical characteristics of OJBT and to analyse simulation results. Measured OJBT data were used for adjusting and calibrating the TCAD simulator from Synopsys' Sentaurus. Gaussian density of states were considered to approximate the carrier's effective density of state in OSCs. The electric-field-dependent mobility Poole–Frenkel mobility model was used to enable the hopping transport of the carriers. We used the constant carrier generation model to compute a constant carrier generation and recombination process.

## GIWAXS analysis

### Evaluation of crystal quality in the in-plane and out-of-plane directions

To evaluate the crystal quality of the differently doped rubrene films for both crystal structures (that is, spherulites and platelets), the (121) reflection ( $Q_{xy} = 1.23 \text{ \AA}^{-1}$  and  $Q_z = 0.23 \text{ \AA}^{-1}$ ) was analysed in the 2D scattering images obtained by GIWAXS measurements ( $Q$  is the scattering vector,  $Q_{xy}$  is the in-plane scattering vector and  $Q_z$  is the out-of-plane scattering vector). Both the in- and out-of-plane directions were analysed to gain information about the crystal quality in the substrate plane and normal to it. We rotated each sample  $360^\circ$  in the substrate plane during the measurements and took individual images every  $1.23^\circ$ .

### Analysis of the in-plane direction

To analyse the in-plane crystal quality, we took single images at specific angles, to minimize the appearance of multiple peaks originating from the

same reflection (caused by different scattering positions on the sample). First, cake segments were extracted from the scattering image ranging from  $Q = 1.15 \text{ \AA}^{-1}$  to  $Q = 1.35 \text{ \AA}^{-1}$  and from  $\chi = 6^\circ$  to  $\chi = 15^\circ$  (where  $\chi$  is the azimuthal angle). The cake segment was then converted into a  $\chi$  versus  $Q$  plot. From this plot, the columns were summed up in an area ranging from  $Q = 1.15 \text{ \AA}^{-1}$  to  $Q = 1.35 \text{ \AA}^{-1}$  and from  $\chi = 6.1^\circ$  to  $\chi = 14.9^\circ$ . Five per cent of the data on each side, horizontally, was used to remove a linear background by fitting.

### Analysis of the out-of-plane direction

To analyse the out-of-plane crystal quality, we averaged the images taken at individual angles. This was possible because the multiple peaks caused by scattering from different positions on the samples result in peaks with their centre aligning on a line from the beam centre. First cake segments were extracted from the scattering image ranging from  $Q = 1.15 \text{ \AA}^{-1}$  to  $Q = 1.35 \text{ \AA}^{-1}$  and from  $\chi = 6^\circ$  to  $\chi = 15^\circ$ . The cake segment was then converted into  $\chi$  versus  $Q$  plot. From this plot, the rows were summed up in an area ranging from  $Q = 1.15 \text{ \AA}^{-1}$  to  $Q = 1.35 \text{ \AA}^{-1}$  and from  $\chi = 6.1^\circ$  to  $\chi = 14.9^\circ$ . Five per cent of the data on each side, vertically, was used to remove a linear background by fitting.

### Peak analysis in the in-plane direction

The resulting spectra were fitted using Gaussian curves and a constant offset. The number of Gaussians used was determined by the goodness of the fit and an estimation of the number of peaks that are distinguishable in the 2D scattering images. The resulting spectra were fitted using a single Gaussian curve with a constant offset.

## Data availability

The data that support the findings of this study are available from <https://opara.zih.tu-dresden.de/xmlui/handle/123456789/2048>.

## References

1. Someya, T., Bao, Z. & Malliaras, G. G. The rise of plastic bioelectronics. *Nature* **540**, 379–385 (2016).
2. Wang, S. et al. Skin electronics from scalable fabrication of an intrinsically stretchable transistor array. *Nature* **555**, 83–88 (2018).
3. Khodagholy, D. et al. In vivo recordings of brain activity using organic transistors. *Nat. Commun.* **4**, 1575 (2013).
4. Sirringhaus, H. 25th Anniversary Article: organic field-effect transistors: the path beyond amorphous silicon. *Adv. Mater.* **26**, 1319–1335 (2014).
5. Klauk, H., Zschieschang, U., Pflaum, J. & Halik, M. Ultralow-power organic complementary circuits. *Nature* **445**, 745–748 (2007).
6. Stutzmann, N., Friend, R. H. & Sirringhaus, H. Self-aligned, vertical-channel, polymer field-effect transistors. *Science* **299**, 1881–1884 (2003).
7. Fratini, S., Nikolka, M., Salleo, A., Schweicher, G. & Sirringhaus, H. Charge transport in high-mobility conjugated polymers and molecular semiconductors. *Nat. Mater.* **19**, 491–502 (2020).
8. Tsurumi, J. et al. Coexistence of ultra-long spin relaxation time and coherent charge transport in organic single-crystal semiconductors. *Nat. Phys.* **13**, 994–998 (2017).
9. Myny, K. The development of flexible integrated circuits based on thin-film transistors. *Nat. Electron.* **1**, 30–39 (2018).
10. Lüssem, B. et al. Doped organic transistors operating in the inversion and depletion regime. *Nat. Commun.* **4**, 2775 (2013).
11. Kleemann, H., Krechan, K., Fischer, A. & Leo, K. A review of vertical organic transistors. *Adv. Funct. Mater.* **30**, 1907113 (2020).

12. Kheradmand-Boroujeni, B. et al. Small-signal measurement technique enabling 40 MHz operation of vertical organic transistors. *Sci. Rep.* **8**, 7643 (2018).
13. Borchert, J. W. et al. Flexible low-voltage high-frequency organic thin-film transistors. *Sci. Adv.* **6**, eaaz5156 (2020).
14. Perinot, A., Giorgio, M., Mattoli, V., Natali, D. & Caironi, M. Organic electronics picks up the pace: mask-less, solution processed organic transistors operating at 160 MHz. *Adv. Sci.* **8**, 2001098 (2021).
15. Guo, E. et al. Vertical organic permeable dual-base transistors for logic circuits. *Nat. Commun.* **11**, 4725 (2020).
16. Klauk, H. Will we see gigahertz organic transistors? *Adv. Electron. Mater.* **4**, 1700474 (2018).
17. Higgins, S. G. et al. Self-aligned organic field-effect transistors on plastic with picofarad overlap capacitances and megahertz operating frequencies. *Appl. Phys. Lett.* **108**, 023302 (2016).
18. Chakraborty, P. S. et al. 130 nm, 0.8 THz fmax, 1.6 V BVCEO SiGe HBTs operating at 4.3 K. *IEEE Electron Device Lett.* **35**, 151–153 (2014).
19. Friend, R. H. et al. Excitons and charges at organic semiconductor heterojunctions. *Faraday Discuss.* **155**, 339–348 (2012).
20. Schwarze, M. et al. Band structure engineering in organic semiconductors. *Science* **352**, 1446–1449 (2016).
21. Burlingame, Q. et al. Centimetre-scale electron diffusion in photoactive organic heterostructures. *Nature* **554**, 77–80 (2018).
22. Sarkar, D. & Halas, N. J. Dember effect in C<sub>60</sub> thin films. *Solid State Commun.* **90**, 261–265 (1994).
23. Lunt, R. R., Giebink, N. C., Belak, A. A., Benziger, J. B. & Forrest, S. R. Exciton diffusion lengths of organic semiconductor thin films

measured by spectrally resolved photoluminescence quenching. *J. Appl. Phys.* **105**, 053711 (2009).

24. Lunt, R. R., Benziger, J. B. & Forrest, S. R. Relationship between crystalline order and exciton diffusion length in molecular organic semiconductors. *Adv. Mater.* **22**, 1233–1236 (2010).
25. Siegmund, B. et al. Exciton diffusion length and charge extraction yield in organic bilayer solar cells. *Adv. Mater.* **29**, 1604424 (2017).
26. Sawatzki, M. F. et al. Doped highly crystalline organic films: toward high-performance organic electronics. *Adv. Sci.* **8**, 2003519 (2021).
27. Yamamura, A. et al. High-speed organic single-crystal transistor responding to very high frequency band. *Adv. Funct. Mater.* **30**, 1909501 (2020).
28. Najafov, H., Lee, B., Zhou, Q., Feldman, L. C. & Podzorov, V. Observation of long-range exciton diffusion in highly ordered organic semiconductors. *Nat. Mater.* **9**, 938–943 (2010).
29. Park, S.-W. et al. Rubrene thin-film transistors with crystalline and amorphous channels. *Appl. Phys. Lett.* **90**, 153512 (2007).
30. Verreet, B., Heremans, P., Stesmans, A. & Rand, B. P. Microcrystalline organic thin-film solar cells. *Adv. Mater.* **25**, 5504–5507 (2013).
31. Fratini, S., Ciuchi, S., Mayou, D., Trambly de Laissardière, G. & Troisi, A. A map of high-mobility molecular semiconductors. *Nat. Mater.* **16**, 998–1002 (2017).
32. Podzorov, V., Menard, E., Rogers, J. A. & Gershenson, M. E. Hall effect in the accumulation layers on the surface of organic semiconductors. *Phys. Rev. Lett.* **95**, 226601 (2005).
33. Dacuña, J. & Salleo, A. Modeling space-charge-limited currents in organic semiconductors: extracting trap density and mobility. *Phys. Rev. B* **84**, 195209 (2011).

34. Fusella, M. A. et al. Use of an underlayer for large area crystallization of rubrene thin films. *Chem. Mater.* **29**, 6666–6673 (2017).
35. Wang, S.-J. et al. Vacuum processed large area doped thin-film crystals: a new approach for high-performance organic electronics. *Mater. Today Phys.* **17**, 100352 (2021).
36. Fusella, M. A. et al. Homoepitaxy of crystalline rubrene thin films. *Nano Lett.* **17**, 3040–3046 (2017).
37. Lee, H. et al. Abrupt heating-induced high-quality crystalline rubrene thin films for organic thin-film transistors. *Org. Electron.* **12**, 1446–1453 (2011).
38. Lee, H., Kim, J., Choi, J. & Cho, S. In situ patterning of high-quality crystalline rubrene thin films for high-resolution patterned organic field-effect transistors. *ACS Nano* **5**, 8352–8356 (2011).
39. Aonuma, M., Oyamada, T., Sasabe, H., Miki, T. & Adachi, C. Material design of hole transport materials capable of thick-film formation in organic light emitting diodes. *Appl. Phys. Lett.* **90**, 183503 (2007).
40. Tietze, M. L. et al. Elementary steps in electrical doping of organic semiconductors. *Nat. Commun.* **9**, 1182 (2018).
41. Bisri, S., Takenobu, T., Takahashi, T. & Iwasa, Y. Electron transport in rubrene single-crystal transistors. *Appl. Phys. Lett.* **96**, 183304 (2010).

## Acknowledgements

We thank A. Hiess and F. Winkler for fabrication of the stencil masks. The GIWAXS experiments were performed at the B11 NCD-SWEET beamline at ALBA Synchrotron with the collaboration of ALBA staff. F.T. and S.M. acknowledge financial support from the German Research Foundation (DFG, MA 3342/6-1) and acknowledge support by the German Excellence Initiative through the Cluster of Excellence EXC 1056 Centre for Advancing Electronics Dresden (cfaed). K.L. acknowledges funding by DFG project Le747/52.

# **Author information**

## **Author notes**

1. These authors contributed equally: Shu-Jen Wang, Michael Sawatzki

## **Authors and Affiliations**

1. Dresden Integrated Center for Applied Physics and Photonic Materials (IAPP), Technische Universität Dresden, Dresden, Germany

Shu-Jen Wang, Michael Sawatzki, Jörn Vahland, Hans Kleemann & Karl Leo

2. NanoP, Technische Hochschule Mittelhessen, University of Applied Science, Gießen, Germany

Ghader Darbandy & Alexander Kloes

3. Center for Advancing Electronics Dresden (cfaed), Technische Universität Dresden, Dresden, Germany

Felix Talnack, Stefan Mannsfeld & Karl Leo

4. ALBA Synchrotron, Barcelona, Spain

Marc Malfois

## **Contributions**

S.-J.W., M.S., H.K. and K.L. designed and planned the experiments. S.-J.W. and M.S. performed the device fabrication and electrical characterization with input from J.V. G.D. and A.K. performed the TCAD simulations. F.T., M.M. and S.M. performed the GIWAXS analysis. H.K. and K.L. supervised the work. All authors discussed the results and contributed to manuscript preparation.

## **Corresponding author**

Correspondence to [Karl Leo](#).

## Ethics declarations

### Competing interests

The authors declare no competing interests.

### Peer review

### Peer review information

*Nature* thanks the anonymous reviewers for their contribution to the peer review of this work. [Peer reviewer reports](#) are available.

### Additional information

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

### Extended data figures and tables

#### [Extended Data Fig. 1 Doped rubrene thin film crystals and their electrical characteristics.](#)

**(a)** Schematic illustration of the crystallization method. **(b)** Polarized microscope images of orthorhombic platelets and spherulites at different doping concentrations (wt.%). **(c)**  $IV$  characteristics of undoped rubrene films in three different crystal phases: Stack consists of 30 nm of undoped seed and 370 nm of undoped bulk film between Au-electrodes (active area of  $100\text{ }\mu\text{m}\times 100\text{ }\mu\text{m}$ ). **(d)**  $IV$  characteristic of orthorhombic spherulite in vertical direction with different concentrations of the p-dopant F6-TCNNQ:Stack consists of 30 nm of undoped seed and 370 nm of doped bulk film between Au-electrodes. **(e)**  $IV$ -curve for different orthorhombic spherulite rubrene crystal thicknesses and SCLC fitting. The  $V^2$ -regime expected from an

SCLC is fitted with orange lines and used to calculate a vertical mobility of  $3.3 \pm 2.5 \text{ cm}^2 \text{V}^{-1} \text{s}^{-1}$  (details given in SI) .

### Extended Data Fig. 2 Morphology of rubrene thin-film seed crystals.

Surface properties measured via AFM of undoped orthorhombic rubrene platelets under different magnifications and growth conditions. (a) crystal grown without the sublayer (30 nm seed only). (b-d) crystal grown with 5 nm of TAPC as sublayer (30 nm seed, 80 nm bulk). (e, f) crystal grown with 5 nm of TAPC as sublayer and 40 nm of Al between seed and bulk (30 nm seed, 80 nm bulk).

### Extended Data Fig. 3 Morphology of rubrene thin-film crystals with doping.

Surface properties measured via AFM of orthorhombic rubrene platelets doped with F6-TCNNQ under different magnifications. (a) crystal grown with 5 nm of TAPC as sublayer and 5 wt.% of F6-TCNNQ (30 nm seed, 80 nm bulk). (b) crystal grown with 5 nm of TAPC as sublayer and 20 wt.% of F6-TCNNQ (30 nm seed, 80 nm bulk).

### Extended Data Fig. 4 X-ray characterization of the doped rubrene thin-film crystals.

(a) Overview of an entire GIWAXS measurement for an orthorhombic platelet film. Structural characterization of thin films. Width of 221-peak from GIWAXS measurements extracted from fit of Gaussian distributions of orthorhombic spherulite (b) and orthorhombic platelet (c) crystals. Inset shows example peak and corresponding fitting. Details regarding the extraction of the peak width are given in the experimental section. The number in round bracket after seed (doping in the seed layer) or bulk (subsequent doping in the bulk film) denotes the doping concentration in wt. % and p/n in square bracket denotes p-type or n-type doping.

### Extended Data Fig. 5 GIWAXS analysis.

**(a)** Peak shape of 121 signal for a spherulite crystal film extracted from GIWAXS measurement. **(b)** Peak shape of 121 signal for a platelet crystal film extracted from GIWAXS measurement. **(c)** Example fit for the in-plane fitting procedure.

### Extended Data Fig. 6 Vertical charge transport analysis.

**(a)** SCLC analysis of charge carrier mobility of orthorhombic spherulite films in vertical direction (p-type doped layers at bottom and top electrode for injection). The SCLC regime was extracted from devices with 400nm and 600nm thickness with eight devices per thickness of varying active area. The error bars denote the standard deviation calculated from multiple devices and different device active areas (the thinner devices show a larger spread). The resulting mobility is calculated from the  $1/L^3$ -dependence of the Mott Gurney law. The uncertainty of the value is based upon the variation measured from the individual devices. The inset shows the SCLC fittings as shown in the Fig. 1e. **(b)**  $IV$  characteristic of orthorhombic platelets crystals in vertical direction with different concentrations of the p-dopant F6-TCNNQ: Stack (inset) consists of 30 nm of undoped seed and 370 nm of doped bulk film between Au-electrodes. Crystals are grown on 5nm TAPC as sublayer.

### Extended Data Fig. 7 Capacitance measurements.

Area normalized capacitance of an individual rubrene-based pinip device at different biasing conditions and varying measurement frequencies. The active area is  $150 \mu\text{m} \times 75 \mu\text{m}$ . The device is fully symmetric and consist of two times 200 nm p-doped, two times 200 nm intrinsic, and 40 nm n-doped rubrene.

### Extended Data Fig. 8 Additional OBJT characterization.

**(a)**  $IV$  measurements of the individual components of the OBJT shown in Fig. 2c. The third, unused electrode is left floating in each of the individual measurements. **(b)** Added current at the output, collector with increased base current for a device with the same stack design as the one shown in Fig. 2d but with a thicker base (50 nm) doped at a higher doping concentration (5

wt.%) for different emitter-collector voltages of  $-12$  and  $-20$ V. (c) Resulting amplification for a device with the same stack design as the one shown in Fig. 2d but with a thicker base (50 nm) doped at a higher doping concentration (5 wt.%) for different emitter-collector voltages of  $-12$  and  $-20$ V.

### **Extended Data Fig. 9 Further thickness and temperature dependent OBJT measurements.**

(a) Resulting amplification of devices with the same stack design as the one shown in Fig. 2d but with a higher doping concentration (5 wt.%) and base width of 10 nm and 50 nm, respectively. As expected, higher doping of the base and thicker base layer reduce amplification. From the change in amplification with base thickness  $W$ , an estimation for the diffusion length can be extracted via  $\beta \propto \coth(W/L_D)$ . As an average from these calculations, a value of 50 nm can be extracted. (b) Temperature dependent differential amplification of devices with the same stack design as the one shown in Fig. 2d. The temperature dependent differential amplification of the device implies an increase in charge diffusion length with temperature which is consistent with a diffusion driven device. The temperature values in the legend are in K.

### **Extended Data Fig. 10 OBJT device layout.**

Schematic cross-section of the fabricated OBJT with 10-Fingers (a) design 1, (b) design 2, and the relevant dimensions of one single finger (c) design 1, (d) design 2.

## **Supplementary information**

### **Supplementary Information**

Supplementary Figs. 1–4 and Table 1.

### **Peer Review File**

## Rights and permissions

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

### Reprints and Permissions

## About this article

### Cite this article

Wang, SJ., Sawatzki, M., Darbandy, G. *et al.* Organic bipolar transistors. *Nature* **606**, 700–705 (2022). <https://doi.org/10.1038/s41586-022-04837-4>

- Received: 09 March 2021
- Accepted: 05 May 2022
- Published: 22 June 2022
- Issue Date: 23 June 2022
- DOI: <https://doi.org/10.1038/s41586-022-04837-4>

### Share this article

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

[Get shareable link](#)

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

[Copy to clipboard](#)

Provided by the Springer Nature SharedIt content-sharing initiative

## **Crystalline order offers access to high speeds for organic transistors**

- Julie Euvrard
- Barry P. Rand

Nature News & Views 22 Jun 2022

---

This article was downloaded by **calibre** from <https://www.nature.com/articles/s41586-022-04837-4>

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

- Article
- Open Access
- [Published: 15 June 2022](#)

# The source of the Black Death in fourteenth-century central Eurasia

- [Maria A. Spyrou](#),
- [Lyazzat Musralina](#),
- [Guido A. Gnechi Ruscone](#),
- [Arthur Kocher](#),
- [Pier-Giorgio Borbone](#),
- [Valeri I. Khartanovich](#),
- [Alexandra Buzhilova](#),
- [Leyla Djansugurova](#),
- [Kirsten I. Bos](#),
- [Denise Kühnert](#),
- [Wolfgang Haak](#),
- [Philip Slavin](#) &
- [Johannes Krause](#)

*Nature* volume **606**, pages 718–724 (2022)

- 192k Accesses
- 1 Citations
- 3363 Altmetric
- [Metrics details](#)

## Abstract

The origin of the medieval Black Death pandemic (ad 1346–1353) has been a topic of continuous investigation because of the pandemic's extensive demographic impact and long-lasting consequences<sup>1,2</sup>. Until now, the most debated archaeological evidence potentially associated with the pandemic's initiation derives from cemeteries located near Lake Issyk-Kul of modern-day Kyrgyzstan<sup>1,3,4,5,6,7,8,9</sup>. These sites are thought to have housed victims of a fourteenth-century epidemic as tombstone inscriptions directly dated to 1338–1339 state 'pestilence' as the cause of death for the buried individuals<sup>9</sup>. Here we report ancient DNA data from seven individuals exhumed from two of these cemeteries, Kara-Djigach and Burana. Our synthesis of archaeological, historical and ancient genomic data shows a clear involvement of the plague bacterium *Yersinia pestis* in this epidemic event. Two reconstructed ancient *Y. pestis* genomes represent a single strain and are identified as the most recent common ancestor of a major diversification commonly associated with the pandemic's emergence, here dated to the first half of the fourteenth century. Comparisons with present-day diversity from *Y. pestis* reservoirs in the extended Tian Shan region support a local emergence of the recovered ancient strain. Through multiple lines of evidence, our data support an early fourteenth-century source of the second plague pandemic in central Eurasia.

## Main

The Black Death, caused by the bacterium *Y. pestis*<sup>10</sup>, was the initial wave of a nearly 500-year-long pandemic termed the second plague pandemic and is one of the largest infectious disease catastrophes in human history<sup>1,11,12</sup>. Estimated to have claimed the lives of up to 60% of the western Eurasian population over its eight-year course<sup>1,12</sup>, the Black Death had a profound demographic and socioeconomic impact in all affected areas, with the European historical record being the most extensively studied resource until now<sup>2,13,14,15</sup>.

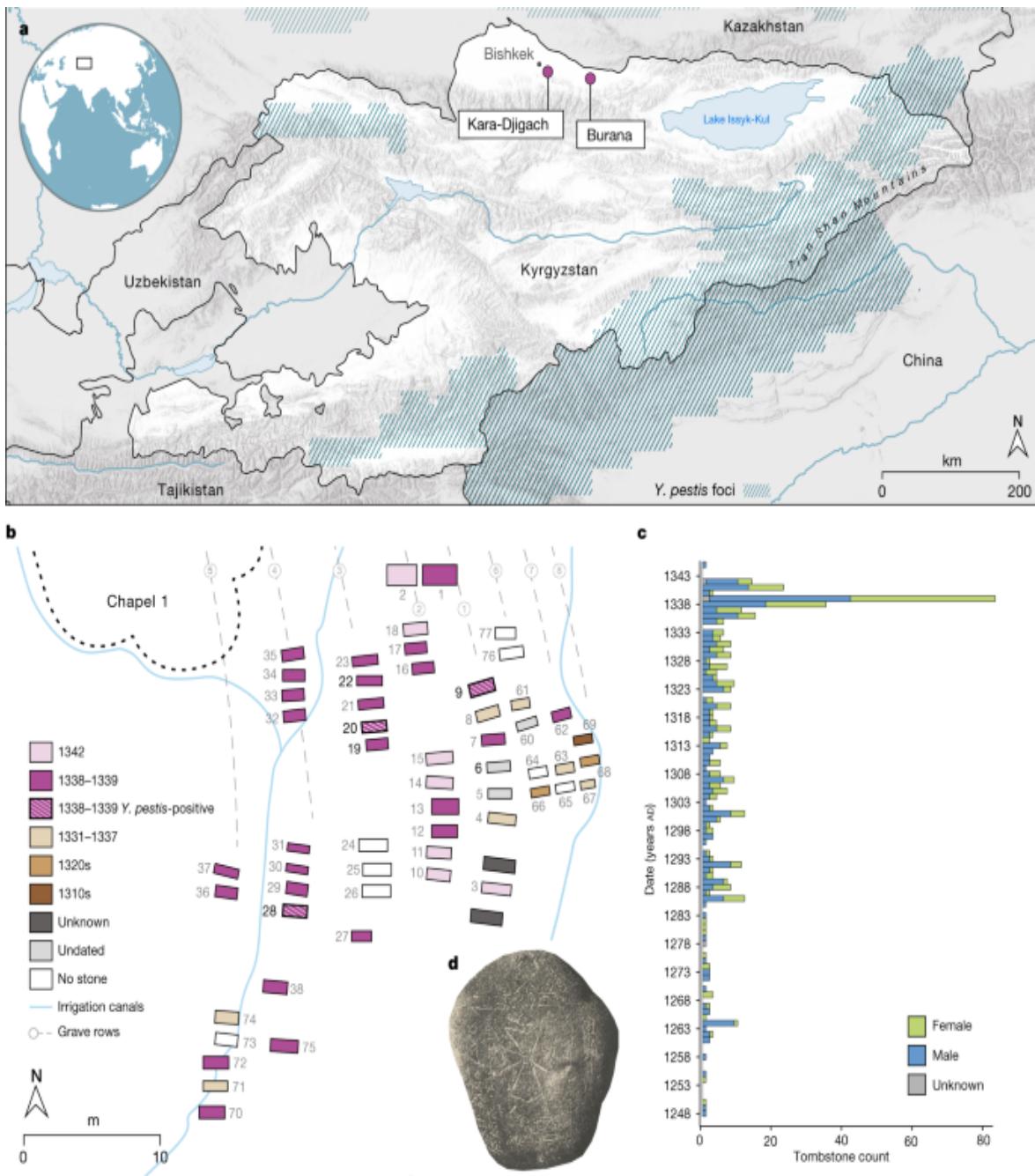
Despite intense multidisciplinary research on this topic, the geographical source of the second plague pandemic remains unclear. Hypotheses based on historical records and modern genomic data have put forward a number of putative source locations ranging from western Eurasia to eastern Asia

(Supplementary Information 1). In recent years, comparisons between ancient and modern *Y. pestis* genomes have shown the Black Death to be associated with a star-like emergence of four major lineages (branches 1, 2, 3 and 4)<sup>16,17</sup>, the descendants of which are dispersed among rodent foci in Eurasia, Africa and the Americas. Although extant lineages that diverged before this event have been identified in central and eastern Eurasia<sup>16,18,19</sup>, complementary ancient DNA (aDNA) data from such regions are lacking. Until now, analyses of the historical record and ancient *Y. pestis* data have largely focused on the pandemic's progression in western Eurasia<sup>12,17,20,21</sup>. Although efforts to expand historical investigations and provide a wider spatiotemporal perspective are under way<sup>9,11,22,23,24,25,26</sup>, the prevailing Eurocentric focus has hampered an identification of the origins of the Second Pandemic.

## A fourteenth-century epidemic in central Eurasia

To explore possible evidence associated with the early history of the second plague pandemic, we investigated the cemeteries of Kara-Djigach and Burana, located in the Chüy Valley near Lake Issyk-Kul of modern-day Kyrgyzstan. Excavations of these cemeteries between 1885 and 1892 revealed a unique archaeological assemblage potentially associated with an epidemic that affected the region during the fourteenth century (Fig. 1 and Supplementary Information 2). On the basis of tombstone inscriptions, these cemeteries showed a disproportionately high number of burials dating between 1338 and 1339, with some inscriptions stating that the cause of death was due to an unspecified pestilence<sup>9,27</sup> (Fig. 1, Extended Data Fig. 1, Supplementary Fig. 1, Supplementary Table 1 and Supplementary Information 2). Given the location, timing and associated demographic pattern, early interpretations considered these characteristics as indicative of a plague epidemic<sup>3,27</sup> and have since triggered a long-lasting debate about the epidemic's association with the onset of the second plague pandemic<sup>1,3,4,5,6,7,8,9,26</sup> (Supplementary Information 2).

**Fig. 1: Description of the investigated fourteenth-century Chüy Valley archaeological sites.**



**a**, Locations of the Kara-Djigach and Burana archaeological sites in modern-day Kyrgyzstan. Regions encompassing *Y. pestis* foci at present are highlighted in blue (as in refs. 18,19). The map was created using QGIS v.3.22.1 (ref. 51) and uses Natural Earth vector map data from <https://www.naturalearthdata.com/>. **b**, Area within the Kara-Djigach cemetery, referred to as ‘Chapel 1’ with the highest concentration of excavated burials dating between 1338 and 1339. Burial dates were

determined on the basis of their associated tombstones (Supplementary Information 2). The site map has been redrawn based on the original created by N. Pantusov in 1885. Individuals from graves 6, 9, 20, 22 and 28 (the numbers in bold) were investigated using aDNA in this study. Burials shown with stripe patterns were associated with individuals BSK001, BSK003 and BSK007, which showed evidence of *Y. pestis* infections. **c**, Annual numbers of tombstones from Kara-Djigach ( $n = 456$ ) and Burana ( $n = 11$ ) (Supplementary Table 1). Dataset updated from ref. 9 (see Supplementary Information 2 for details). **d**, Tombstone from the Kara-Djigach cemetery with legible pestilence-associated inscription. The inscription is translated as ‘In the Year 1649 [=ad 1338], and it was the Year of the Tiger, in Turkic Bars. This is the tomb of the believer Sanmaq. [He] died of pestilence [=mawtānā]’. For a tracing of the inscription, see Extended Data Fig. 1.

To better understand the contexts of Kara-Djigach and Burana, we translated and analysed surviving archival information from their excavations (Supplementary Information 2 and Supplementary Figs. 1–4). Furthermore, we generated human genomic data from 7 individuals (5 from Kara-Djigach and 2 from Burana) through a hybridization capture of approximately 1.24 million ancestry-informative single-nucleotide polymorphisms (SNPs)<sup>28</sup>, which resulted in 4 individuals with sufficient genomic coverage for population genetic analyses (>30,000 SNPs). Using principal component analysis and ancestry modelling, we found these individuals to be falling broadly within the variability of ancient and present-day populations from central Eurasia. However, precise connections could not be determined given the scarcity of contemporaneous human genomic data from this region (Supplementary Information 3, Supplementary Fig. 5 and Supplementary Tables 2–5). On the basis of the available tombstone inscriptions, burial artefacts, coin hoards and historical records, we found that the Chüy Valley housed ethnically diverse communities that relied on trade and maintained connections with several regions across Eurasia (Supplementary Information 2). Such links may have contributed to the spread of infectious diseases to and from this region during the fourteenth century.

## Ancient pathogen DNA screening

To investigate traces of ancient pathogen DNA that could explain the cause of the suspected epidemic, shotgun metagenomic data generated from all seven individuals were taxonomically classified using the HOPS pipeline<sup>29</sup> (Supplementary Table 6). Of those, three individuals exhumed from the Kara-Djigach cemetery (BSK001, BSK003 and BSK007) displayed potential evidence of ancient *Y. pestis* DNA (Supplementary Table 7) as well as low edit distances in reads mapping against the CO92 reference genome, and the presence of chemical alterations characteristic of aDNA (Supplementary Fig. 6 and Supplementary Table 8). As such, the respective DNA libraries were subjected to whole-genome *Y. pestis* capture ([Methods](#)).

## The ancestor of a fourteenth-century polytomy

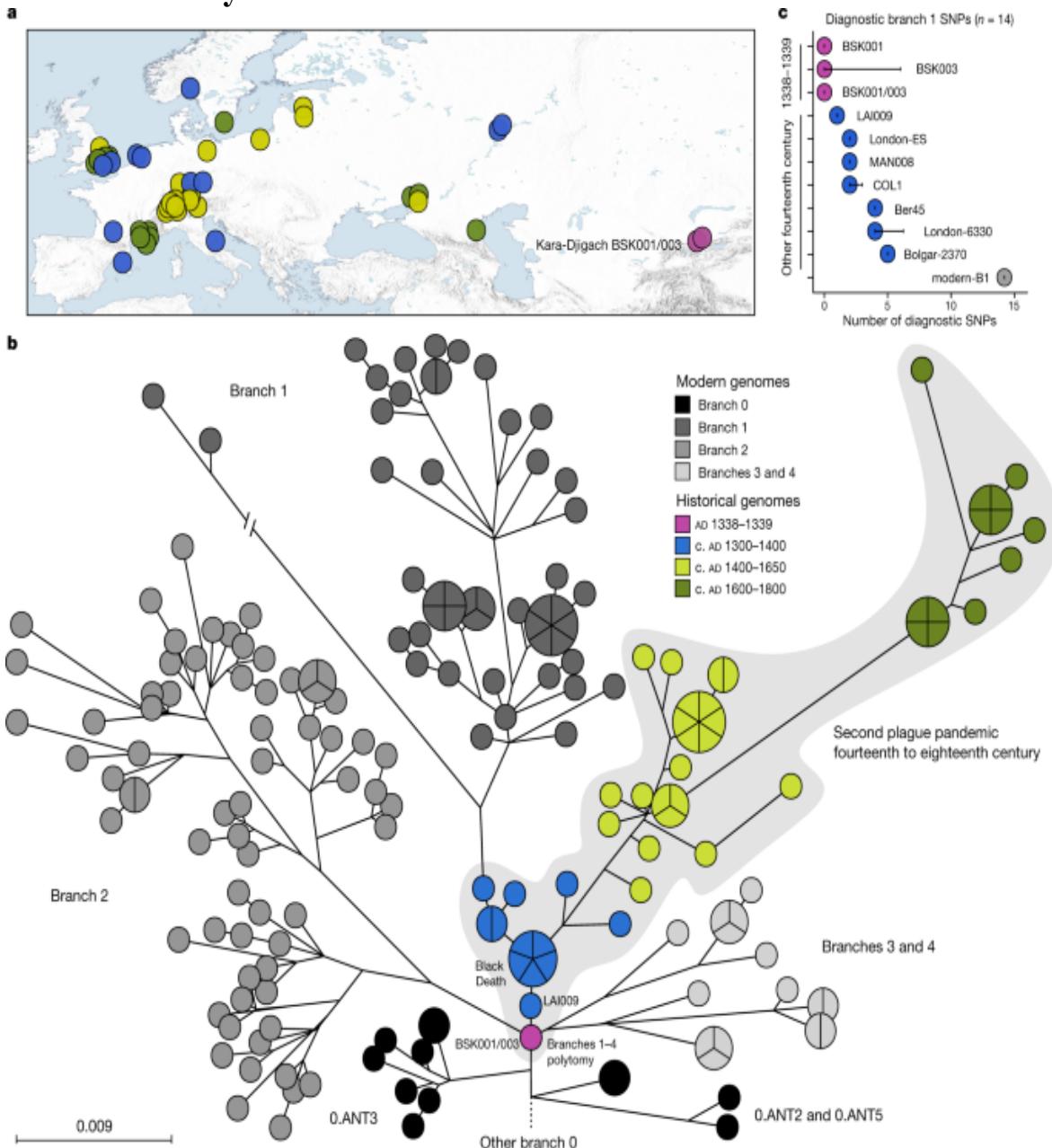
Whole-genome *Y. pestis* capture yielded 6.7-fold and 2.8-fold average coverage for BSK001 and BSK003, respectively. Coverage across all three *Y. pestis* plasmids ranged from 24.7-fold to 4.7-fold (Supplementary Tables 9 and 10). For BSK007, genomic coverage was lower, approximately 0.13-fold, resulting from poorer aDNA preservation that was also reflected in the shotgun screening and human DNA enrichment data (Supplementary Tables 2, 3 and 8). Nevertheless, this sample was considered a true *Y. pestis*-positive because of the even distribution of mapping reads against the CO92 reference chromosome and the presence of aDNA-associated damage (Extended Data Figs. 2 and 3 and Supplementary Tables 9–11). Furthermore, a metagenomic classification of BSK007 reads aligning to the pCD1, pMT1 and pPCP1 plasmids identified >99% as *Y. pestis*-specific (Extended Data Fig. 3).

To evaluate whether the higher-coverage *Y. pestis* genomes BSK001 and BSK003 represented distinct bacterial strains, we compared their SNP profiles. To limit variant calls deriving from environmental contamination, particularly given the high amounts of multi-allelic sites identified in both genomes (Supplementary Fig. 7), we performed a taxonomy-informed metagenomic filtering using MALT ([Methods](#) and Supplementary Table 11). We identified 20 sites differing between BSK001 and BSK003, all of which are unique variants in the lower-coverage BSK003 (Supplementary Table 12). On the basis of previously defined authenticity criteria<sup>30,31</sup> ([Methods](#)), all such variants were consistent with residual exogenous contamination,

suggesting that the two genomes were probably identical. Recovery of identical strains from both individuals is consistent with published evidence showing low diversity in *Y. pestis* genomes isolated from single epidemic contexts<sup>10,17,20,21,32</sup>. On the basis of their associated tombstones, BSK001, BSK003 and BSK007 were buried during the epidemic year 1338–1339 (Fig. 1 and Supplementary Information 2) and our data further support a *Y. pestis* involvement in this event.

We performed a comparative SNP analysis between the Kara-Djigach genomes and previously published historical and currently circulating *Y. pestis* diversity (Fig. 2a, Supplementary Tables 13–15). For this, BSK001 and BSK003 were combined (BSK001/003) to achieve an increased genomic resolution (combined coverage of 9.5-fold; Supplementary Table 9). Our analysis revealed one SNP unique to BSK001/003 when compared against 203 modern and 46 historical *Y. pestis* chromosomal genomes (Extended Data Fig. 4 and Supplementary Tables 16 and 17). This SNP was found in a region with persistent multi-allelic sites; therefore, it is considered artefactual<sup>31</sup> (Supplementary Fig. 8). Consistent with previous research on the evolutionary history of *Y. pestis*<sup>16</sup>, our inferred maximum likelihood phylogeny exhibited five major branches, designated 0, 1, 2, 3 and 4, with published Second Pandemic genomes being associated with branch 1 (Fig. 2b). The placement of BSK001/003 is ancestral to all published fourteenth-century genomes from western Eurasia (Fig. 2b and Extended Data Fig. 5), separated by one SNP from LAI009, an isolate from the Volga region in eastern Europe<sup>17</sup>, and by two SNPs from five genetically identical Black-Death-associated genomes from southern, central and northern Europe<sup>17,21</sup>. Specifically, BSK001/003 is positioned on a node previously designated N07 (ref. 16), which preceded the multifurcation of branches 1–4. To evaluate whether missing data affected the accuracy of our phylogenetic placements, we investigated all BSK001 and BSK003 variant calls for shared positions with lineages deriving from the N07 node and those directly preceding it. BSK001/003 carries the ancestral state in all covered diagnostic SNPs defining branches 1–4 and 0.ANT3, which is the closest related branch 0 lineage to BSK001/003, as well as the derived state in all positions leading from 0.ANT3 to N07 (Fig. 2c, Extended Data Fig. 6 and Supplementary Table 18). At our current resolution, we conclude that BSK001/003 represents the direct progenitor of the branch 1–4 polytomy.

**Fig. 2: Comparisons between BSK001/003 and published *Y. pestis* genomic diversity.**



**a**, Map of all historical *Y. pestis* genomes used in the present study ( $n = 48$ ). The colours represent different genome ages on a scale between 1300 and 1800, as depicted in **b**. The colour scale is maintained across all panels of this figure. To aid visibility in overlapping symbols, a jitter option was implemented for plotting genomes on the map. The map was created with QGIS v.3.22.1 (ref. 51) and uses Natural Earth vector map data from <https://www.naturalearthdata.com/>. **b**, *Y. pestis* maximum likelihood

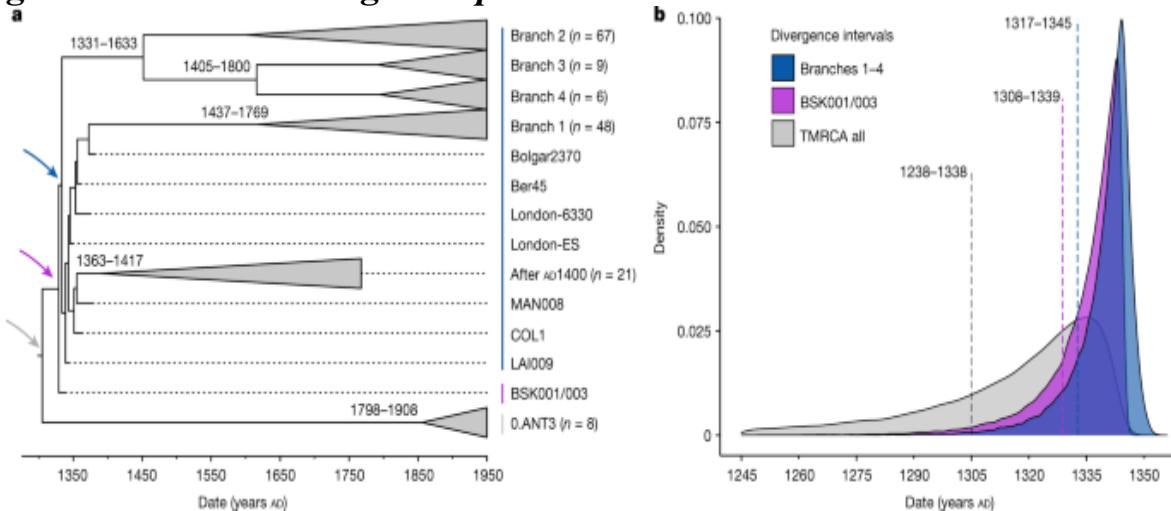
phylogeny based on 2,960 SNPs, visualized using GrapeTree<sup>50</sup>. The depicted portion of the phylogeny contains the closest related lineages to BSK001/003. (For a fully labelled tree, see Extended Data Fig. 5). The colours of published historical strains are consistent with a. The scale denotes the number of substitutions per genomic site. c, Abundance of diagnostic SNP sharing in fourteenth-century *Y. pestis* genomes. The number of diagnostic SNPs ( $n$ ) shared between all modern genomes on branch 1, and therefore defining this branch, were retrieved from a comparative SNP table of 203 modern *Y. pestis* genomes. SNP sharing was assessed by determining the allele status of each diagnostic position according to a threefold SNP calling threshold. The error bars denote the degree of missing data ( $n$ ) in the respective ancient genome. Refer to Extended Data Fig. 6 and Supplementary Table 18 for an overview of diagnostic SNP sharing on different phylogenetic branches.

## Divergence time for the branch 1–4 polytomy

The polytomy of branches 1–4 is a major event in the evolution of *Y. pestis* given its association with the Black Death<sup>9,26,33</sup> and the rich genetic diversity that emerged from it<sup>16</sup> (Fig. 2b). Estimates on the timing of this diversification have so far yielded wide ranges spanning from the tenth to the fourteenth centuries<sup>16,34</sup>. Recently, a narrower time frame was proposed that placed this emergence in the early thirteenth century, more than 100 years before the Black Death<sup>22,26</sup>. As BSK001/003 represents the common ancestor of branches 1–4, we used this genome from 1338 to 1339 to construct a time-calibrated phylogeny and re-estimate an age range for this diversification with BEAST2 (Supplementary Figs. 9 and 10 and Supplementary Table 19). After evaluating a number of demographic models (Supplementary Table 20), our resulting estimates based on the coalescent Bayesian skyline model revealed overlapping ages for the divergence of BSK001/003 (95% highest posterior density (HPD): 1308–1339), as well as for that of branch 1 from branches 2–4 (95% HPD: 1317–1345) (Fig. 3). As BEAST2 only infers bifurcating trees, we also used TreeTime<sup>35</sup> to infer a time-calibrated phylogeny that can retain polytomies. Consistent with our estimates above, we inferred a 1316–1340 date for the split time of branches 1–4 (Supplementary Fig. 11), although we caution that this method does not

account for age uncertainties in ancient genomes. Taken together, the present results support an age range spanning the first half of the fourteenth century for the timing of the branch 1–4 polytomy.

**Fig. 3: Molecular dating of *Y. pestis* branches 1–4.**



**a**, Maximum clade credibility time-calibrated phylogenetic tree. The tree is based on 167 genomes (historical and modern) and was estimated using the coalescent skyline tree prior and a log-normal relaxed clock. Collapsed branches contain modern and ancient isolates dating after AD 1400 (post-Black Death). The coloured arrows mark the nodes, for which equivalent posterior age distributions are shown in **b**. The estimated divergence dates (95% HPD intervals) of modern branches are shown on each corresponding node. **b**, Estimated posterior distributions based on the coalescent Bayesian skyline tree prior for the divergence of *Y. pestis* branches 1–4 (blue), for the estimated divergence of BSK001/003 (purple) and for the entire dataset used for this analysis (time to the most recent common ancestor of branches 1–4 and 0.ANT3, shown in grey). The dotted lines indicate mean posterior estimates and are annotated with the corresponding 95% HPD intervals.

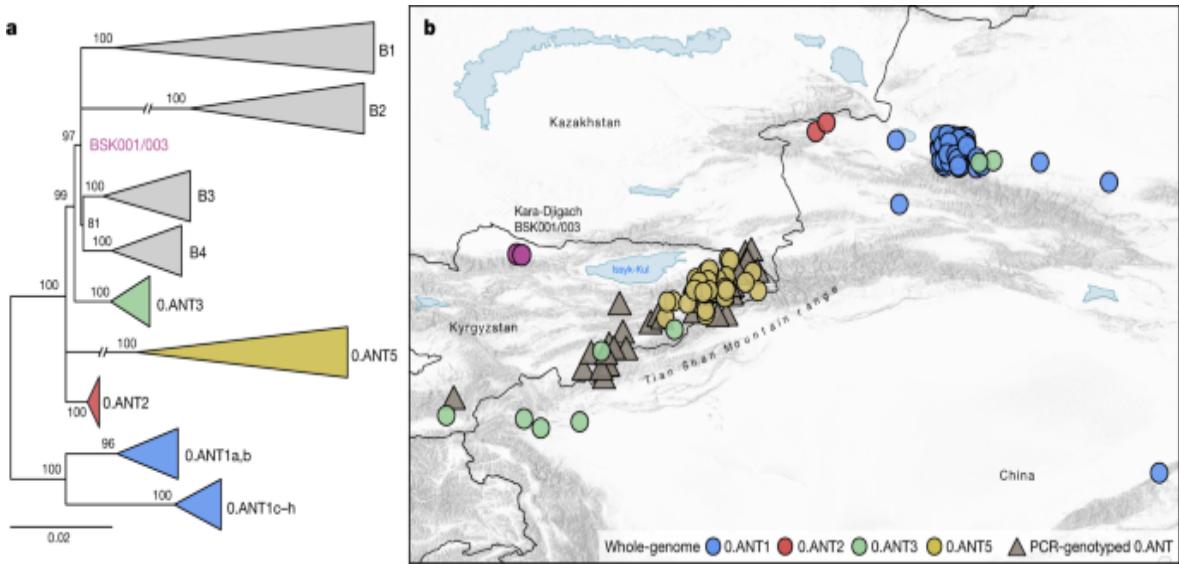
Furthermore, to quantify the proportion of present-day *Y. pestis* genetic diversity that emerged from this polytomy, we computed mean pairwise distances (MPDs) and Faith's phylogenetic diversity (FPD) indices in 203 genomes comprising our entire modern dataset, as well as 130 genomes comprising branches 1–4 ([Methods](#)). In our dataset, 64% (130 out of 203) of modern *Y. pestis* strains belonged to branches 1–4, reflecting the high worldwide frequency known for these lineages<sup>[16,36,37](#)</sup>. We estimate that

branches 1–4 represent approximately 40% of the overall phylogenetic diversity within present-day *Y. pestis* based on our full dataset (MPD ratio: 41%; 95% percentile interval (PI): 35.3–46.4; FPD ratio: 35.9%; 95% PI: 31.6–39.5). This value is marginally reduced after equalizing the number of genomes in branches 1–4 and branch 0 (MPD ratio: 36.8%; 95% PI: 32.0–41.9; FPD ratio: 33.9%; 95% PI: 29.4–37.7) (Extended Data Fig. 7). Given that the known history of *Y. pestis* reaches back at least 5,000 years<sup>38</sup>, it is notable that a substantial fraction of its surviving genetic diversity accumulated since the fourteenth century.

## Plague reservoirs in the Tian Shan area

To address existing hypotheses on the Black Death's geographical origins (Supplementary Information 1), we investigated the possibility of a local emergence versus an introduction of the BSK001/003 strain into the Chüy Valley from a different area. For this, we assessed the geographical distribution of the most closely related ancestral branching lineages to BSK001/003 and identified 164 present-day 0.ANT strains with record of their isolation locations (Supplementary Table 21). Consistent with previous interpretations<sup>9,18,26</sup>, we found that all such strains were retrieved from foci in eastern Kazakhstan, eastern Kyrgyzstan and the Xinjiang Uygur Autonomous Region of northwestern China (Fig. 4 and Extended Data Fig. 8). Although we cannot exclude a different geographical range for these lineages in the past, the current data are consistent with a local emergence of BSK001/003 within the extended Tian Shan region. Intriguingly, the oldest recovered genome associated with 0.ANT was also identified in the Tian Shan region (third century AD)<sup>39</sup> and forms part of an extinct clade that caused the first plague pandemic (sixth to eighth centuries AD)<sup>30</sup>. As noted previously<sup>18,26,33,40</sup>, most extant 0.ANT strains have been isolated from marmots and their ectoparasites known to be the primary *Y. pestis* reservoirs in these areas (Supplementary Table 21). Therefore, such species could represent possible candidates for the spillover that led to the second plague pandemic.

**Fig. 4: Geographical isolation locations of modern 0.ANT lineages.**



**a**, Maximum likelihood phylogenetic tree, based on 2,441 genome-wide variant positions. The tree was constructed to indicate the genetic relationships between available 0.ANT genomes depicted on the map and BSK001/003. Modern branches were collapsed to enhance tree clarity (see Extended Data Fig. 8 for a full tree). **b**, Map depicting the geographical isolation locations of 0.ANT strains (Supplementary Table 21), which belong to the closest ancestral branching lineages to the Kara-Djigach strain. The map includes both whole-genome data (further specified as 0.ANT lineages 1, 2, 3 and 5) and PCR-genotyped isolates that are broadly defined as 0.ANT, belonging to any of the 4 lineages. For strains in which exact geographical coordinates were unavailable, locations were approximated according to their associated plague reservoirs. To aid visibility in overlapping symbols, a jitter option was implemented for plotting objects on the map. The map was created with QGIS v.3.22.1 (ref. 51) and uses Natural Earth vector map data from <https://www.naturalearthdata.com/>.

## Discussion

The power of ancient metagenomics lies in its potential to provide direct evidence for testing long-standing historical hypotheses and reveal phylogeographical patterns of microbial diversity through time<sup>41</sup>. One such debate concerns the events that triggered the second plague pandemic, as well as the time and place of its emergence. Recently, an analysis of historical, genetic and ecological data led to the suggestion that the

emergence of *Y. pestis* branches 1–4 occurred more than a century before the beginning of the Black Death. According to the proposed model, this initial diversification was mediated by people and was linked with territorial expansions of the Mongol Empire across Eurasia during the early thirteenth century<sup>22,26,42</sup>. By contrast, we present ancient *Y. pestis* data from central Eurasia that support a fourteenth-century emergence; therefore, earlier outbreak attributions remain to be explored. At present, the narrow-focused sampling chosen for this study does not allow for an assessment of the spread of the BSK001/003 strain. Previous studies have shown that *Y. pestis* can disseminate rapidly without accumulation of genetic diversity<sup>17,21</sup>, thus potentiating a contemporaneous presence of the same strain across a large geographical range. Nevertheless, the known range of extant plague foci associated with lineages ancestral to BSK001/003 provide support for its emergence in central Eurasia and possibly in the extended Tian Shan region. Although the dynamics that triggered the bacterium's emergence in this region are unknown, previous studies showed that environmental factors, such as natural disasters and sudden changes in temperature and precipitation can have an impact on *Y. pestis* host ecologies and, as a result, can trigger outbreaks in human populations<sup>43,44,45,46</sup>. Although we have no evidence to infer such connections with the Kara-Djigach epidemic, we envision that our precise 1338–1339 date will serve as a reference point for future environmental, archaeological and historical research focusing on the events that caused a *Y. pestis* introduction into human populations and precipitated the second plague pandemic.

The onset of the Black Death has been conventionally associated with outbreaks that occurred around the Black Sea region in 1346 (refs. <sup>1,47</sup>), eight years after the Kara-Djigach epidemic. At present, the exact means through which *Y. pestis* reached western Eurasia are unknown, primarily due to large pre-existing uncertainties around the historical and ecological contexts of this process. Previous research suggested that both warfare and/or trade networks were some of the main contributors in the spread of *Y. pestis*<sup>21,22,26,47,48</sup>. Yet, related studies have so far either focused on military expeditions that were arguably unrelated to initial outbreaks<sup>47</sup> or others that occurred long before the mid-fourteenth century<sup>22,26</sup>. Moreover, even though preliminary analyses exist to support an involvement of Eurasian-wide trade routes in the spread of the disease<sup>48</sup>, their systematic exploration has so far

been conducted only for restricted areas of western Eurasia<sup>21,47</sup>. The placement of the Kara-Djigach settlement in proximity to trans-Asian networks<sup>9,49</sup>, as well as the diverse toponymic evidence and artefacts identified at the site (Supplementary Information 2) lend support to scenarios implicating trade in *Y. pestis* dissemination. Therefore, an investigation of early-to-mid-fourteenth-century connections across Asia, interpreted alongside genomic evidence, will be important for disentangling the bacterium's westward dispersals.

Past and present experiences have demonstrated that reconciling the source of a pandemic is a complex task that cannot be accomplished by a single research discipline. Although the ancient *Y. pestis* genomes reported in this Article offer biological evidence to settle an old debate, it is the unique historical and archaeological contexts that define our study's scope and importance. As such, we envision that future synergies will continue to reveal important insights for a detailed reconstruction of the processes that triggered the second plague pandemic.

## Methods

### **Sampling, DNA extraction, partial uracil DNA glycosylase library preparation and sequencing**

We obtained permission from the Kunstkamera, Peter the Great Museum of Anthropology and Ethnography in St Petersburg for the sampling and ancient DNA analysis of 7 tooth specimens, excavated between 1885 and 1892 from the medieval cemeteries of Kara-Djigach and Burana (Supplementary Information 2). No statistical methods were used to predetermine the number of samples used in this study. All laboratory procedures were carried out in the dedicated aDNA facilities of the Max Planck Institute for the Science of Human History and the Max Planck Institute for Evolutionary Anthropology. The detailed procedures used for tooth sampling can be found in ref. 52. In brief, teeth were sectioned in the dentin–enamel junction using an electric saw with a diamond blade. After tooth sectioning, approximately 50 mg of powder was removed from the surface of the pulp chamber of each tooth using rounded dental drill bits.

The recovered tooth powder was used for DNA extractions using a previously established protocol optimized for the recovery of short fragments of DNA<sup>53</sup>. The exact steps and modifications of the procedure used have been made available in ref. <sup>54</sup>. In brief, the tooth powder was incubated overnight (12–16 h) at 37 °C in 1 ml of DNA lysis buffer containing EDTA (0.45 M, pH 8.0) and proteinase K (0.25 mg ml<sup>-1</sup>). After incubation, DNA binding and isolation was performed using a custom GuHCl-based binding buffer and purification using High Pure Viral Nucleic Acid Large Volume Kit (Roche). Finally, DNA was eluted in 100 µl of Tris-EDTA-Tween containing Tris-HCl (10 mM), EDTA (1 mM, pH 8.0) and Tween-20 (0.05%). For procedure monitoring, extraction blanks and positive extraction controls were included throughout the laboratory processing steps.

All DNA extracts were converted into one-to-two double-stranded DNA libraries for Illumina sequencing, using 25 µl of input extract per library with an initial partial uracil DNA glycosylase (UDG) and endonuclease VIII treatment (USER enzyme; New England Biolabs) according to established protocols<sup>55,56</sup>. The detailed library preparation procedure, including the blunt-end repair, adaptor ligation and adaptor fill-in reaction steps can be found in ref. <sup>57</sup>. After library preparation, each library was quantified using a quantitative PCR system (LightCycler 96 Instrument) using the IS7 and IS8 primers<sup>55</sup>. For multiplex sequencing, we performed double indexing of all libraries using previously published procedures<sup>58</sup>, outlined in detail in ref. <sup>59</sup>. A combination of unique index primers containing 8 base pair (bp) identifiers were assigned to each library. To aid amplification efficiency, libraries were then split into multiple PCR reactions for the indexing step based on their initial IS7/IS8 quantification. The number of indexing PCR reactions performed for each library was determined so that every reaction was assigned an input of no more than  $1.5 \times 10^{10}$  DNA copies. Each reaction was set up using the Pfu Turbo Cx Hotstart DNA Polymerase (Agilent Technologies) and was run for 10 cycles using the following conditions: initial denaturation at 95 °C for 2 min followed by a cycling of 95 °C for 30 s, 58 °C for 30 s and 72 °C for 1 min, as well as a final elongation step at 72 °C for 10 min. All PCR products were purified using the MinElute DNA Purification Kit (QIAGEN), with some modifications to the manufacturer's protocol<sup>59</sup>. Finally, all indexing PCR products were qPCR-quantified

(LightCycler 96 Instrument) using the IS5 and IS6 primer combination<sup>58,59</sup>. To avoid heteroduplex formation, indexed libraries were amplified to  $10^{13}$  DNA copies per reaction with the Herculase II Fusion DNA Polymerase (Agilent Technologies) and quantified using a 4200 Agilent TapeStation Instrument using a D1000 ScreenTape system (Agilent Technologies). Libraries were diluted to 10 nM and pooled equimolarly for sequencing. We performed shotgun DNA sequencing on an Illumina HiSeq 4000 platform using a 76-cycle kit ( $1 \times 76 + 8 + 8$  cycles).

## Shotgun next-generation sequencing read processing and metagenomic screening

After demultiplexing, raw shotgun sequenced reads were preprocessed in the EAGER pipeline v.1.92.58 using AdapterRemoval v.2.2.0 (ref. <sup>60</sup>), which was used to remove Illumina adaptors (minimum overlap of 1 bp), as well as for read filtering according to sequencing quality (minimum base quality of 20) and length (retaining reads  $\geq 30$  bp). Subsequently, all datasets were screened for the presence of pathogen DNA traces using the metagenomic pipeline HOPS<sup>29</sup>. First, preprocessed reads were aligned against a custom RefSeq database<sup>61</sup> (November 2017) containing all complete bacterial and viral genome assemblies, a subset of eukaryotic pathogen assemblies and the GRCh38 human reference genome. Genome assemblies that contained the word ‘unknown’ were removed from the database, retaining a total of 15,361 entries. The database retained a number of *Yersinia* species entries: *Yersinia alcovae* ( $n = 1$ ), *Yersinia aleksiciae* ( $n = 1$ ), *Yersinia enterocolitica* ( $n = 16$ ), *Yersinia entomophaga* ( $n = 1$ ), *Yersinia frederiksenii* ( $n = 3$ ), *Yersinia intermedia* ( $n = 1$ ), *Yersinia kristensenii* ( $n = 2$ ), *Y. pestis* ( $n = 39$ ), *Yersinia phage* ( $n = 17$ ), *Yersinia pseudotuberculosis* ( $n = 13$ ), *Yersinia rohdei* ( $n = 1$ ), *Yersinia ruckeri* ( $n = 4$ ), *Yersinia similis* ( $n = 1$ ) and *Yersinia* sp. FDA-ARGOS ( $n = 1$ ). MALT v0.4<sup>62</sup> was run using the following parameters: -id 90 -lcaID 90 -m BlastN -at SemiGlobal -topMalt 1 -sup 1 -mq 100 -verboseMalt 1 -memoryMode load -additionalMaltParameters. The resulting alignment files were post-processed with MALTEExtract for a qualitative assessment against a predefined list of 356 target taxonomic entries ([https://github.com/rhuebler/HOPS/blob/external/Resources/default\\_list.txt](https://github.com/rhuebler/HOPS/blob/external/Resources/default_list.txt)). Specifically, reads were assessed according to their edit distance against a

specific pathogen sequence in the database and the potential occurrence of mismatches that could signify the presence of aDNA damage<sup>29</sup>. In cases in which both parameters were met, the corresponding pathogen alignment was considered a strong candidate. Preprocessed reads were mapped against the *Y. pestis* CO92 (NC\_003143.1) and human (*hg19*) reference genomes with the Burrows–Wheeler Aligner (BWA). Mapping parameters were set to 0.01 for the edit distance (-n) and seed length was disabled (-l 9999).

Subsequently, we used SAMtools v.1.3 (ref. <sup>63</sup>) to remove reads with mapping quality lower than 37 (for CO92) or 30 (for *hg19*); PCR duplicates were removed with MarkDuplicates v1.140 (<http://broadinstitute.github.io/picard/>). Finally, patterns of aDNA damage were assessed with mapDamage v.2.0 (ref. <sup>64</sup>).

## Single-stranded DNA library preparation and hybridization capture

For specimens BSK001 and BSK003, extra single-stranded DNA libraries were constructed from an input DNA extract of 30 µl. We performed library preparation at the Max Planck Institute for Evolutionary Anthropology using an automated protocol that is publicly available<sup>65</sup>. Single-stranded and double-stranded libraries from individuals BSK001, BSK003 and BSK007 were enriched using DNA probes covering the whole *Y. pestis* genome, as well as 1.24 million genome-wide SNP sites of the human genome<sup>66,67</sup>. For capture preparation, all libraries were amplified for the necessary number of PCR cycles to achieve 1–2 µg of input DNA. PCR reactions were carried out using the Herculase II Fusion DNA Polymerase. They were then purified using the MinElute DNA Purification Kit and eluted in EB elution buffer containing 0.05% Tween 20. Finally, library concentrations (ng µl<sup>-1</sup>) were quantified using a NanoDrop spectrophotometer (Thermo Fisher Scientific). For the in-solution *Y. pestis* captures, the probe set design was based on a set of publicly available modern genomes, specifically the *Y. pestis* CO92 chromosome (NC\_003143.1), CO92 plasmid pMT1 (NC\_003134.1), CO92 plasmid pCD1 (NC\_003131.1), KIM10 chromosome (NC\_004088.1), Pestoides F chromosome (NC\_009381.1) and the *Y. pseudotuberculosis* IP32953 chromosome (NC\_006155.1). For the in-solution human DNA captures, the probe set design was created to target 1,237,207 variants across the genome that are informative for studying the genetic history of

worldwide human populations<sup>28,67</sup>. Both human DNA and *Y. pestis* hybridization captures were carried out for two rounds as described previously<sup>28,69,68,67,66</sup>, in which partially UDG-treated libraries from the same individual were pooled in equimolar ratios for capture and single-stranded libraries were captured separately.

## Post-capture *Y. pestis* data processing

After *Y. pestis* whole-genome capture, libraries were sequenced on a HiSeq 4000 platform ( $1 \times 76 + 8 + 8$  cycles or  $2 \times 76 + 8 + 8$  cycles) at a depth of approximately 11–27 million raw reads. The preprocessing of raw demultiplexed reads was carried out as described in the ‘Shotgun next-generation sequencing read processing and metagenomic screening’ section. At this stage, the datasets produced from partially UDG-treated libraries from the same individual were pooled and terminal bases were trimmed using fastx\_trimmer (FASTX Toolkit 0.0.14, [http://hannonlab.cshl.edu/fastx\\_toolkit/](http://hannonlab.cshl.edu/fastx_toolkit/)) to avoid damaged site interference with SNP calling during further processing. The following steps for read mapping, PCR duplicate removal and aDNA damage calculation were carried out in the EAGER pipeline<sup>70</sup>. We performed read mapping with BWA v.0.7.12 against the *Y. pestis* CO92 reference genome (NC\_003143.1). For the pooled and trimmed partial UDG-treated libraries, BWA parameters were set to 0.1 for the edit distance (-n) and seed length was disabled (-l 9999). Given that the single-stranded libraries constructed for this study retained aDNA-associated damage, the BWA parameters were set to 0.01 for the edit distance (-n) to allow for an increased number of mismatches that could derive from deamination; seed length was disabled (-l 9999). We performed read mapping against the plasmids using the same parameters against a concatenated reference of all three *Y. pestis* plasmids (pMT1: NC\_003134.1; pPCP1: NC\_003132.1; and pCD1: NC\_003131.1), masking the problematic pPCP1 region between nucleotides 3000 and 4200 that was shown to have high similarity to expression vectors used in laboratory reagents<sup>71</sup>. SAMtools v.1.3 (ref. 63) was used to remove all reads with mapping quality lower than 37 (-q), whereas MarkDuplicates was used to remove PCR duplicates. Deamination patterns associated with aDNA damage were retrieved with mapDamage v.2.0 (ref. 64). We used MALT<sup>62</sup> for a taxonomic classification of mapped reads, to attempt a retention of

reads that are more likely to be endogenous *Y. pestis*. MALT was run against the same database as described in the section ‘Shotgun next-generation sequencing read processing and metagenomic screening’, using the following parameters: -m BlastN -at SemiGlobal -top 1 -sup 1 -mq 100 -memoryMode load -ssc -sps. The minimum percentage identity parameter was set to default (-id 0.0), as opposed to a 90% identity filter used for running HOPS<sup>29</sup>, to avoid any reference bias that might arise from the removal of endogenous reads with a higher number of mismatches. After run completion, to retain the maximum number of reads accounting for the naive lowest common ancestor algorithm, we extracted reads that were assigned to the *Yersinia* genus node or summarized under the *Y. pseudotuberculosis* complex node. Reads were extracted in FASTA format from MEGAN v.6.4.12 (ref. <sup>72</sup>). Subsequently, FASTA files were converted into FASTQ format with the reformat.sh script in BBMap from the BBtools suite (version 38.86, <https://sourceforge.net/projects/bbmap/>). FASTQ files were then remapped against the CO92 reference genome using the same parameters as described previously in this section. For single-stranded libraries, mapDamage v.2.0 (ref. <sup>64</sup>) was used to rescale quality scores in read positions at which potential deamination-associated mismatches to the reference were identified. Subsequently, BAM files corresponding to the same individual were concatenated after mapping quality filtering and PCR duplicate removal. We performed concatenation using the SAMtools ‘merge’ command and with the AddOrReplaceReadGroups tool in Picard (<http://broadinstitute.github.io/picard/>) for assigning a single read group to all reads in each new file.

## SNP calling, heterozygosity estimates and SNP filtering

Variant calling was carried out for BSK001 and BSK003, both before and after MALT<sup>62</sup> filtering using the UnifiedGenotyper in the Genome Analysis Toolkit (GATK) v.3.5 (ref. <sup>73</sup>). GATK was run using the EMIT\_ALL\_SITES option, which produced a call for every position on the chromosomal CO92 reference genome. The resulting genomic profiles of BSK001 and BSK003 were compared against a set of 233 modern and 46 historical *Y. pestis* genomes, as well as against the *Y. pseudotuberculosis* reference genome IP32953 (NC\_006155.1), using the Java tool MultiVCFAnalyzer v.0.85 (<https://github.com/alexherbig/MultiVCFAnalyzer>). MultiVCFAnalyzer

v.0.85 was run with the following parameters. SNPs were called at a minimum coverage of threefold and in cases of heterozygous positions, calls were made at a 90% minimum support threshold. In addition, SNPs were called at a minimum genotyping quality of 30. Furthermore, previously defined non-core and repetitive regions, as well as regions containing homoplasies, ribosomal RNAs, transfer-messenger RNAs and transfer RNAs were excluded from comparative SNP calling<sup>16,32</sup>. A set of 6,567 total variant sites were identified in the present dataset.

To investigate the extent of possible exogenous contamination within the BSK001 and BSK003 datasets, we estimated the number of ambiguous heterozygous variants beyond the SNP calling threshold. For this, MultiVCFAnalyzer v.0.85 (ref. 74) was used to generate an SNP table of alternative allele frequencies ranging between 10 and 90%. The results were then used to create ‘heterozygosity’ histogram plots of the estimated frequencies in R v.3.6.1 (ref. 75). Heterozygosity plots were created both before and after MALT filtering (see ‘Post-capture *Y. pestis* data processing’) to investigate whether taxonomy-informed filtering could aid the elimination of contaminant sequences in the investigated datasets (Supplementary Fig. 7).

An SNP table created with MultiVCFAnalyzer v.0.85, containing all variant positions across the present dataset, was filtered to identify SNP differences between the BSK001 and BSK003 genomes. The identified differences ( $n = 20$ ) were then evaluated with the Java tool SNP\_Evaluation<sup>30</sup> (build date 13 August 2018; [https://github.com/andreasKroepelin/SNP\\_Evaluation](https://github.com/andreasKroepelin/SNP_Evaluation)). The variant table and the VCF files of each genome were used as input for SNP\_Evaluation. Furthermore, each identified private variant was evaluated within a 50 bp window and was considered ‘true’ when fulfilling the following criteria established in studies published previously<sup>17,21,30,76</sup>: (1) no multi-allelic sites were permitted within the evaluated window unless they were consistent with aDNA deamination (signified as spurious C-to-T or G-to-A substitutions); (2) the evaluated SNP position itself was not consistent with aDNA damage (no bases overlapping the SNP were downscaled by mapDamage v.2.0 (ref. 64)); (3) no gaps in genomic coverage were identified in the evaluated window; (4) reads overlapping the SNP sites

showed specificity to the *Y. pseudotuberculosis* complex when screened with BLASTn (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>).

Finally, to gain phylogenetic resolution, the BSK001 and BSK003 *Y. pestis* datasets were concatenated. We performed concatenation of BAM files, MALT<sup>62</sup> filtering and aDNA damage rescaling (with mapDamage v.2.0 (ref. 64)) as described in the section ‘Post-capture *Y. pestis* data processing’.

Moreover, the dataset was included in the comparative SNP analysis using MultiVCFAnalyzer v.0.85 (ref. 74) as described above. Finally, unique SNPs were evaluated with SNP\_Evaluation<sup>30</sup> according to the four criteria listed above.

## Phylogenetic reconstruction and diversity estimations

Phylogenetic analysis was used to explore 233 *Y. pestis* genomes as part of the modern comparative dataset. An SNP alignment produced by MultiVCFAnalyzer v.0.85 (ref. 74) was used to construct a phylogenetic tree in MEGA7, using the maximum parsimony approach with 95% partial deletion (6,032 SNPs). Of the 233 modern *Y. pestis* genomes in the current dataset, 30 displayed extensive private branch lengths (Supplementary Fig. 12). Such an effect in bacterial phylogenies could result either from true biological diversity or from technical artefacts associated with false SNP incorporation during computational genome reconstruction. Although we cannot exclude the presence of several strains with exceedingly higher mutation rates in the current dataset, previous studies showed that modern *Y. pestis* strains with ‘mutator’ profiles are uncommon<sup>16,36</sup>. In this study, 27 out of 30 genomes that showed disparities in their private SNP counts compared to the rest of the dataset, were derived from assemblies for which the quality of SNP calls could not be evaluated (raw data unavailable). Because potential mis-assemblies or false-positive SNP calls can affect evolutionary inferences and diversity estimations, these genomes were excluded from further analyses. Therefore, we performed phylogenetic analysis using a subset of 203 modern *Y. pestis* genomes (Supplementary Table 13). The list of excluded genomes is as follows: 2.MED1\_139 (ref. 19), 2.MED1\_A-1809 (ref. 18), 2.MED1\_A-1825 (ref. 19), 2.MED1\_A-1920 (ref. 19), 2.MED0\_C-627 (ref. 19), 2.MED1\_M-1484 (ref. 19), 2.MED1\_M-519 (ref. 19), 0.ANT5\_A-1691 (ref. 18), 0.ANT5\_A-1836 (ref. 18), 0.PE2\_C-

678 (ref. 77), 0.PE2\_C-370 (ref. 77), 0.PE2\_C-700 (ref. 77), 0.PE2\_C-746 (ref. 77), 0.PE2\_C-535 (ref. 77), 0.PE2\_C-824 (ref. 77), 0.PE2\_C-712 (ref. 77), 0.PE2b\_G8786 (ref. 16), 0.PE4\_I-3446 (ref. 78), 0.PE4\_I-3517 (ref. 78), 0.PE4t\_A-1815 (ref. 18), 0.PE4\_I-3447 (ref. 78), 0.PE4\_I-3518 (ref. 78), 0.PE4\_I-3443 (ref. 78), 0.PE4\_I-3442 (ref. 78), 0.PE4\_I-3519 (ref. 78), 0.PE4\_I-3516 (ref. 78), 0.PE4\_I-3515 (ref. 78), 0.PE4\_Microtus91001 (ref. 79), 0.PE5\_I-2238 (ref. 80) and 0.PE7b\_620024 (ref. 16).

A genome-wide SNP alignment consisting of 203 modern-day and 48 historical *Y. pestis* genomes (Supplementary Table 14), as well as the *Y. pseudotuberculosis* IP32953 genome, was used as input to construct a maximum likelihood phylogeny including 2,960 SNPs and up to 4% missing data. We performed phylogenetic analysis with RAxML<sup>81</sup> v.8.2.9 using the generalized time-reversible (GTR) substitution model with 4 gamma rate categories. Finally, 1,000 bootstrap replicates were used to estimate node support for the resulting tree topology. After run completion, the maximum likelihood phylogenies were visualized with FigTree v.1.4.4 (<http://tree.bio.ed.ac.uk/software/figtree/>) and GrapeTree (v1.5.0)<sup>50</sup>.

To estimate the proportion of modern *Y. pestis* diversity descending from BSK001/003, we used the R package picante v1.8.2<sup>82</sup> to compute the MPD and FPD<sup>83</sup> from the reconstructed maximum likelihood substitution tree. Measures made on a subset of the tree corresponding to the subclade descending from BSK001/003 (branches 1–4) were compared to that of the complete *Y. pestis* phylogeny. In both cases, only modern strains were included in the calculation. We used a bootstrapping approach to assess the sensitivity of our results with regard to sampling and phylogenetic uncertainty<sup>84</sup>. For each of the 1,000 RAxML bootstrap trees, we randomly resampled modern strains with replacement and only kept branches of the tree corresponding to the sampled strains. Diversity measures were performed for each of the obtained resampled bootstrap trees, from which median estimates and 95% percentile intervals were derived.

To assess the potential impact of uneven sampling among branches (branches 1–4 contained 130 modern strains whereas branch 0 contained only 73), we repeated the same analysis but adding an initial step intended to equalize the number of genomes in both parts of the tree. We subsampled

branches 1–4 to the same number of strains as in branch 0 using sequence clustering in branches 1–4 to obtain representative subsamples. We performed hierarchical clustering based on pairwise phylogenetic distances (derived from the maximum likelihood phylogenetic tree) and the resulting tree was cut to define 73 clusters (functions `hclust`<sup>85</sup> and `cutree` in R v.4.0.3). For each bootstrap tree, clusters were randomly downsampled to one strain, resulting in an equal number of strains between branch 1–4 and branch 0. Resampling with replacement was then applied as previously to each of the downsampled trees before computing diversity measures.

## Plasmid SNP analysis

To investigate possible genetic variation among the plasmids of historical genomes, we performed read mapping of BSK001, BSK003 and BSK001/003 with BWA as well as SNP calling with GATK v.3.5 as described in the above section 'SNP calling, heterozygosity estimates and SNP filtering' against each of the three *Y. pestis* plasmids (pMT1: NC\_003134.1; pPCP1: NC\_003132.1; and pCD1: NC\_003131.1). We then performed comparative SNP calling using MultiVCFAnalyzer v0.85 (ref. <sup>74</sup>) against a set of 46 historical *Y. pestis* genomes as well as the modern reference strains CO92, KIM5 and 0.PE4-Microtus91001. Variants were filtered in individual genomes using SNP\_Evaluation according to previously defined criteria (see the 'SNP calling, heterozygosity estimates and SNP filtering' section). In the present dataset, we identified ten variants in pCD1, eight in pMT1 and two in pPCP1 (Supplementary Table <sup>15</sup>).

## Time-calibrated phylogenetic analysis

To estimate the timing for the divergence of *Y. pestis* branches 1–4 using the BSK001/003 genomes as a new calibration point, we used a dataset comprising all modern genomes from branches 1–4 used for phylogenetic analysis ( $n = 130$ ), genomes of the ancestral branching lineage 0.ANT3 ( $n = 8$ ) and all 29 historical (fourteenth–eighteenth century) genomes in our dataset representing unique genotypes. In cases of identical genomes, the highest coverage genome was chosen for this analysis. We applied a molecular clock test using a maximum likelihood method in MEGA7 (ref. <sup>86</sup>), using a GTR substitution model in which differences in evolutionary

rates among sites were estimated using a discrete gamma distribution with four rate categories. On the basis of this molecular clock test, the null hypothesis of equal evolutionary rates across tested phylogenetic branches was rejected, which is consistent with previous studies showing substitution rate variation across *Y. pestis* lineages<sup>16,17</sup>. Therefore, a log-normal relaxed clock model was used for all subsequent molecular dating analyses.

For the molecular dating analysis, we used the Bayesian statistical framework BEAST2 v.6.6 (ref. <sup>87</sup>). The ages of all ancient isolates were used as calibration points to construct a time-calibrated phylogeny with their radiocarbon or archaeological context age ranges set as uniform priors (see Supplementary Table <sup>19</sup> for all used age ranges). The ages of all modern isolates were set to 0 years before the present. We tested a number of coalescent tree priors such as the coalescent constant size, Bayesian skyline<sup>88</sup> and exponential population models, all of which have been used or tested in previous ancient pathogen genomic studies<sup>17,89,90,91</sup>. We also tested the birth–death skyline tree prior, which has gained traction in recent years<sup>91,92,93</sup> because it can account for epidemiological variables and can model sampling disparities through time<sup>94</sup>. Moreover, we used jModelTest v.2.1.10 (ref. <sup>95</sup>) to identify the substitution model of best fit for our dataset. The indicated transversion model was implemented in BEAUTi by using a GTR model (4 gamma rate categories) and the AG substitution rate parameter fixed to 1.0 (as indicated previously<sup>93</sup>). All tree priors were used in combination with a log-normal relaxed clock rate with a uniform prior distribution ranging between  $1 \times 10^{-3}$  and  $1 \times 10^{-6}$  substitutions per site per year for the SNP alignment (1,405 sites after a 95% partial deletion), corresponding to a range of  $3 \times 10^{-7}$  to  $3 \times 10^{-10}$  across the entire genome, which is within the range of previous estimates<sup>17</sup>. As part of the phylogenetic topology set-up, all branch 1–4 genomes (ancient plus modern) as well as the 0.ANT3 lineage were constrained to be independent monophyletic clades. For the constant population size and exponential population tree priors, all other parameters were set to default. For the coalescent skyline tree prior, a Jeffreys prior distribution ( $1/x$ ) was used for the population sizes and a dimension of 5 was used to permit variations in the group and population sizes through time, with an upper bound of 380,000 for the effective population size (default). Moreover, for the birth–death skyline tree prior, we used a uniform prior for the rate to become non-

infectious that ranged between 0.03 and 70, to account for possible infectious periods ranging from 30 years (lifelong infections in rodent reservoirs<sup>96,97</sup>) to 5 days (average infectious period for bubonic plague<sup>98</sup>). We used a prior beta distribution with mean = 0.1 (alpha = 10.0, beta = 90.0) for the sampling probability  $\rho$  at time 0 and a uniform distribution ranging between 0 and 0.1 for the sampling proportion  $s$ . For the latter, two shifts were allowed through time. Finally, the reproductive number  $R$  was allowed to vary between 0 and 4.0 using a long normal prior distribution of median = 1.0 and s.d. = 0.7, which is within the range of previous estimates for bubonic and pneumonic plague during medieval epidemics<sup>98</sup>.

The suitability of all tree priors was evaluated using path sampling as implemented in the model selection package of BEAST2 v.6.6. Path sampling was run in 50 steps, with 20 million states as the chain length for each step. The resulting log-marginal likelihoods favoured with ‘strong support’<sup>99</sup> the coalescent skyline model for the present analysis (log Bayes factor= 8.35 when compared against the second best model) (Supplementary Table 20). Therefore, the coalescent skyline model was chosen for further analysis. To evaluate the temporal signal in the present dataset, we used TempEst v.1.5.3 to estimate the root-to-tip distance against specimen ages in a linear regression analysis<sup>100</sup>. For TempEst, we used a maximum parsimony tree computed in MEGA7 (ref. 86) in NEXUS format. Moreover, we used the midpoint of the archaeological or radiocarbon date ranges for all ancient genomes as tip dates. All modern genome ages were set to 0 years before the present. The resulting correlation coefficient  $r$  (0.39) and  $R^2$  (0.16) values supported the existence of a temporal signal in the present dataset. Furthermore, we used the BETS approach<sup>101</sup> for a temporal signal assessment that takes into account all analysis parameters. BETS compares the (log)-marginal likelihood estimations produced from an isochronous model (all sampling dates set to 0 years before the present) against a heterochronous model (including real sampling times). As previously, path sampling was run in 50 steps with 20 million states as the chain length for each step. The estimated (log)-Bayes factor of 129.33 was in strong support of the heterochronous model; therefore, it indicated the presence of a temporal signal in the present dataset.

For the molecular dating analysis using a coalescent skyline model set-up, we performed Markov chain Monte Carlo sampling using 2 independent chains of 300–400 million states each. After completion, runs were combined using LogCombiner v.2.6.7 and convergence was evaluated using Tracer v.1.6 (<http://tree.bio.ed.ac.uk/software/tracer/>) ensuring that the effective sample sizes were greater than 200 for each estimated posterior distribution after a 10% burn-in. Maximum clade credibility trees were constructed using TreeAnnotator in the BEAST2 v.6.6 package<sup>87</sup> with a 10% burn-in and were then visualized in FigTree v.1.4.4. In parallel with the molecular dating analysis, we performed a sampling from the prior analysis to test for possible overfitting of the prior to the data. We performed Markov chain Monte Carlo sampling for 2 independent chains of 600 million states each. After run completion, runs were combined and convergence was evaluated after a 30% burn-in. The results indicate that the posterior distributions of the uncorrelated log-normal relaxed clock and the time to the most recent common ancestor estimates are not concordant with those obtained when using a data-informed analysis (Supplementary Fig. 13).

Because most Bayesian phylogenetic frameworks (such as BEAST2) are based on bifurcating trees and hence are poor at resolving multifurcating nodes, we complemented our approach by using TreeTime v.0.8.4 (ref. 35) to infer a time-calibrated phylogeny using a maximum likelihood approach. TreeTime has been shown to resolve polytomies in a way that is consistent with specimen tip dates. We generated a rooted maximum likelihood phylogeny using RAxML (Supplementary Fig. 10) from the same SNP alignment as the one used for BEAST2 (95% partial deletion). The maximum likelihood tree was then used as input for TreeTime, which was run using all known sampling dates for modern genomes and the midpoint of the age range for the ancient genomes (Supplementary Table 22). TreeTime was run using the Kingman coalescent tree prior with the skyline setting. An appropriate substitution model was chosen for the data using the -gtr infer option. The time-scaled phylogeny was inferred using an uncorrelated relaxed clock and with the branch length optimization, keep-root and keep-polytomies options. Moreover, the divergence time intervals were estimated from the highest likelihood tree using the -confidence option. Analyses were run using a maximum number of 500 and 1,000 iterations (maximum

number of iterations option) and produced consistent outputs. The resulting time tree can be found in Supplementary Fig. [11](#).

## Reporting summary

Further information on research design is available in the [Nature Research Reporting Summary](#) linked to this paper.

## Data availability

The raw sequence data produced in this study, the *Y. pestis* aligned reads after metagenomic filtering and the human aligned reads are available through the European Nucleotide Archive under accession no. [PRJEB46734](#). More data are available in the [Supplementary Information](#).

## Code availability

No specialized custom code was used for this study. All software used for the data analyses in this study is publicly available.

## References

1. Benedictow, O. J. *The Black Death, 1346–1353: the Complete History* (Boydell & Brewer, 2004).
2. Clark, G. *A Farewell to Alms: A Brief Economic History of the World* Vol. 25 (Princeton Univ. Press, 2008).
3. Pollitzer, R. *Plague* (World Health Organization, 1954).
4. Norris, J. East or west? The geographic origin of the Black Death. *Bull. Hist. Med.* **51**, 1–24 (1977).
5. Dols, M. W. *The Black Death in the Middle East* (Princeton Univ. Press, 1979).

6. McNeill, W. H. *Plagues and Peoples* (Anchor, 1976).
7. Campbell, B. M. *The Great Transition. Climate, Disease and Society in the Late-Medieval World* (Cambridge Univ. Press, 2016).
8. Stewart, J. *Nestorian Missionary Enterprise: The Story of a Church on Fire* (T. & T. Clark, 1928).
9. Slavin, P. Death by the lake: mortality crisis in early fourteenth-century Central Asia. *J. Interdiscip. Hist.* **50**, 59–90 (2019).
10. Bos, K. I. et al. A draft genome of *Yersinia pestis* from victims of the Black Death. *Nature* **478**, 506–510 (2011).
11. Varlik, N. *Plague and Empire in the Early Modern Mediterranean World* (Cambridge Univ. Press, 2015).
12. Benedictow, O. J. *The Complete History of the Black Death* (Boydell & Brewer, 2021).
13. Alfani, G. & Murphy, T. E. Plague and lethal epidemics in the pre-industrial world. *J. Econ. Hist.* **77**, 314–343 (2017).
14. Herlihy, D. *The Black Death and the Transformation of the West* (Harvard Univ. Press, 1997).
15. Pamuk, S. The Black Death and the origins of the ‘Great Divergence’ across Europe, 1300–1600. *Eur. Rev. Econ. Hist.* **11**, 289–317 (2007).
16. Cui, Y. et al. Historical variations in mutation rate in an epidemic pathogen, *Yersinia pestis*. *Proc. Natl Acad. Sci. USA* **110**, 577–582 (2013).
17. Spyrou, M. A. et al. Phylogeography of the second plague pandemic revealed through analysis of historical *Yersinia pestis* genomes. *Nat. Commun.* **10**, 4470 (2019).
18. Eroshenko, G. A. et al. *Yersinia pestis* strains of ancient phylogenetic branch 0.ANT are widely spread in the high-mountain plague foci of

Kyrgyzstan. *PLoS ONE* **12**, e0187230 (2017).

19. Kutyrev, V. V. et al. Phylogeny and classification of *Yersinia pestis* through the lens of strains from the plague foci of Commonwealth of Independent States. *Front. Microbiol.* **9**, 1106 (2018).
20. Guellil, M. et al. A genomic and historical synthesis of plague in 18th century Eurasia. *Proc. Natl Acad. Sci. USA* **117**, 28328–28335 (2020).
21. Namouchi, A. et al. Integrative approach using *Yersinia pestis* genomes to revisit the historical landscape of plague during the Medieval Period. *Proc. Natl Acad. Sci. USA* **115**, E11790–E11797 (2018).
22. Hymes, R. Epilogue: a hypothesis on the East Asian beginnings of the *Yersinia pestis* polytomy. *Mediev. Globe* **1**, 285–308 (2015).
23. Chouin, G. Reflections on plague in African history (14th–19th c.). *Afriques* <https://doi.org/10.4000/afriques.2228> (2018).
24. Fazlinejad, A. & Ahmadi, F. The Black Death in Iran, according to Iranian historical accounts from the fourteenth through fifteenth centuries. *J. Persianate Stud.* **11**, 56–71 (2018).
25. Borsch, S. & Sabraa, T. Refugees of the Black Death: quantifying rural migration for plague and other environmental disasters. *Ann. Demogr. Hist.* **134**, 63–93 (2017).
26. Green, M. H. The four Black Deaths. *Am. Hist. Rev.* **125**, 1601–1631 (2020).
27. Chwolson, D. *Syrisch-nestorianische Grabinschriften aus Semirjetschie, neue Folge* (Commissionnaires de l'Académie Impériale des Sciences, 1897).
28. Fu, Q. et al. An early modern human from Romania with a recent Neanderthal ancestor. *Nature* **524**, 216–219 (2015).
29. Hübler, R. et al. HOPS: automated detection and authentication of pathogen DNA in archaeological remains. *Genome Biol.* **20**, 280

(2019).

30. Keller, M. et al. Ancient *Yersinia pestis* genomes from across Western Europe reveal early diversification during the First Pandemic (541–750). *Proc. Natl Acad. Sci. USA* **116**, 12363–12372 (2019).
31. Warinner, C. et al. A robust framework for microbial archaeology. *Annu. Rev. Genom. Hum. Genet.* **18**, 321–356 (2017).
32. Bos, K. I. et al. Eighteenth century *Yersinia pestis* genomes reveal the long-term persistence of an historical plague focus. *eLife* **5**, e12994 (2016).
33. Green, M. H. How a microbe becomes a pandemic: a new story of the Black Death. *Lancet Microbe* **1**, e311–e312 (2020).
34. Spyrou, M. A. et al. Analysis of 3800-year-old *Yersinia pestis* genomes suggests Bronze Age origin for bubonic plague. *Nat. Commun.* **9**, 2234 (2018).
35. Sagulenko, P., Puller, V. & Neher, R. A. TreeTime: maximum-likelihood phylodynamic analysis. *Virus Evol.* **4**, vex042 (2018).
36. Morelli, G. et al. *Yersinia pestis* genome sequencing identifies patterns of global phylogenetic diversity. *Nat. Genet.* **42**, 1140–1143 (2010).
37. Zhou, Z. et al. The Enterobase user’s guide, with case studies on *Salmonella* transmissions, *Yersinia pestis* phylogeny, and *Escherichia* core genomic diversity. *Genome Res.* **30**, 138–152 (2020).
38. Susat, J. et al. A 5,000-year-old hunter–gatherer already plagued by *Yersinia pestis*. *Cell Rep.* **35**, 109278 (2021).
39. de Barros Damgaard, P. et al. 137 ancient human genomes from across the Eurasian steppes. *Nature* **557**, 369–374 (2018).
40. Anisimov, A. P., Lindler, L. E. & Pier, G. B. Intraspecific diversity of *Yersinia pestis*. *Clin. Microbiol. Rev.* **17**, 434–464 (2004).

41. Spyrou, M. A., Bos, K. I., Herbig, A. & Krause, J. Ancient pathogen genomics as an emerging tool for infectious disease research. *Nat. Rev. Genet.* **20**, 323–340 (2019).
42. Fancy, N. A. G. & Green, M. H. Plague and the fall of Baghdad (1258). *Med. Hist.* **65**, 157–177 (2021).
43. McCormick, M. Rats, communications, and plague: toward an ecological history. *J. Interdiscip. Hist.* **34**, 1–25 (2003).
44. Pourhossein, B., Esmaeili, S., Gyuranecz, M. & Mostafavi, E. Tularemia and plague survey in rodents in an earthquake zone in southeastern Iran. *Epidemiol. Health* **37**, e2015050 (2015).
45. Dennis, D. T. et al. *Plague Manual: Epidemiology, Distribution, Surveillance and Control* (World Health Organization, 1999).
46. Schmid, B. V. et al. Climate-driven introduction of the Black Death and successive plague reintroductions into Europe. *Proc. Natl Acad. Sci. USA* **112**, 3020–3025 (2015).
47. Barker, H. Laying the corpses to rest: grain, embargoes, and *Yersinia pestis* in the Black Sea, 1346–48. *Speculum* **96**, 97–126 (2021).
48. Gómez, J. M. & Verdú, M. Network theory may explain the vulnerability of medieval human settlements to the Black Death pandemic. *Sci. Rep.* **7**, 43467 (2017).
49. Liu, X. *The Silk Road in World History* (Oxford Univ. Press, 2010).
50. Zhou, Z. et al. GrapeTree: visualization of core genomic relationships among 100,000 bacterial pathogens. *Genome Res.* **28**, 1395–1404 (2018).
51. QGIS Geographic Information System <http://www.qgis.org> (QGIS, 2022).
52. Neumann, G. U., Andrades Valtuena, A., Fellows Yates, J. A., Stahl, R. & Brandt, G. Tooth sampling from the inner pulp chamber for ancient

DNA extraction V.2. *protocols.io*  
<https://doi.org/10.17504/protocols.io.bqebmtan> (2020).

53. Dabney, J. et al. Complete mitochondrial genome sequence of a Middle Pleistocene cave bear reconstructed from ultrashort DNA fragments. *Proc. Natl Acad. Sci. USA* **110**, 15758–15763 (2013).
54. Velsko, I., Skourtanioti, E. & Brandt, G. Ancient DNA extraction from skeletal material. *protocols.io*  
<https://doi.org/10.17504/protocols.io.baksicwe> (2020).
55. Meyer, M. & Kircher, M. Illumina sequencing library preparation for highly multiplexed target capture and sequencing. *Cold Spring Harb. Protoc.* **2010**, pdb.prot5448 (2010).
56. Rohland, N., Harney, E., Mallick, S., Nordenfelt, S. & Reich, D. Partial uracil-DNA-glycosylase treatment for screening of ancient DNA. *Philos. Trans. R. Soc. Lond. B* **370**, 20130624 (2015).
57. Aron, F., Neumann, G. U. & Brandt, G. Half-UDG treated double-stranded ancient DNA library preparation for illumina sequencing. *protocols.io* <https://doi.org/10.17504/protocols.io.bmh6k39e> (2020).
58. Kircher, M., Sawyer, S. & Meyer, M. Double indexing overcomes inaccuracies in multiplex sequencing on the Illumina platform. *Nucleic Acids Res.* **40**, e3 (2012).
59. Stahl, R. et al. Illumina double-stranded DNA dual indexing for ancient DNA. *protocols.io* <https://doi.org/10.17504/protocols.io.bakticwn> (2020).
60. Schubert, M., Lindgreen, S. & Orlando, L. AdapterRemoval v2: rapid adapter trimming, identification, and read merging. *BMC Res. Notes* **9**, 88 (2016).
61. O’Leary, N. A. et al. Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res.* **44**, D733–D745 (2016).

62. Vågene, Å. J. et al. *Salmonella enterica* genomes from victims of a major sixteenth-century epidemic in Mexico. *Nat. Ecol. Evol.* **2**, 520–528 (2018).
63. Li, H. et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
64. Jónsson, H., Ginolhac, A., Schubert, M., Johnson, P.-L. F. & Orlando, L. mapDamage2.0: fast approximate Bayesian estimates of ancient DNA damage parameters. *Bioinformatics* **29**, 1682–1684 (2013).
65. Gansauge, M.-T., Aximu-Petri, A., Nagel, S. & Meyer, M. Manual and automated preparation of single-stranded DNA libraries for the sequencing of DNA from ancient biological remains and other sources of highly degraded DNA. *Nat. Protoc.* **15**, 2279–2300 (2020).
66. Andrades Valtueña, A. et al. The Stone Age plague and its persistence in Eurasia. *Curr. Biol.* **27**, 3683–3691.e8 (2017).
67. Mathieson, I. et al. Genome-wide patterns of selection in 230 ancient Eurasians. *Nature* **528**, 499–503 (2015).
68. Fu, Q. et al. DNA analysis of an early modern human from Tianyuan Cave, China. *Proc. Natl Acad. Sci. USA* **110**, 2223–2227 (2013).
69. Haak, W. et al. Massive migration from the steppe was a source for Indo-European languages in Europe. *Nature* **522**, 207–211 (2015).
70. Peltzer, A. et al. EAGER: efficient ancient genome reconstruction. *Genome Biol.* **17**, 60 (2016).
71. Schuenemann, V. J. et al. Targeted enrichment of ancient pathogens yielding the pPCP1 plasmid of *Yersinia pestis* from victims of the Black Death. *Proc. Natl Acad. Sci. USA* **108**, E746–E752 (2011).
72. Huson, D. H. et al. MEGAN Community Edition - interactive exploration and analysis of large-scale microbiome sequencing data. *PLoS Comput. Biol.* **12**, e1004957 (2016).

73. DePristo, M. A. et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Genet.* **43**, 491–498 (2011).
74. Bos, K. I. et al. Pre-Columbian mycobacterial genomes reveal seals as a source of New World human tuberculosis. *Nature* **514**, 494–497 (2014).
75. R Core Team. *R: A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, 2019).
76. Feldman, M. et al. A high-coverage *Yersinia pestis* genome from a sixth-century Justinianic plague victim. *Mol. Biol. Evol.* **33**, 2911–2923 (2016).
77. Kislichkina, A. A. et al. Eight whole-genome assemblies of *Yersinia pestis* subsp. *microtus* bv. *Caucasica* isolated from the common vole (*Microtus arvalis*) plague focus in Dagestan, Russia. *Genome Announc.* **5**, e00847-17 (2017).
78. Kislichkina, A. A. et al. Nine whole-genome assemblies of *Yersinia pestis* subsp. *microtus* bv. *Altaica* strains isolated from the Altai Mountain natural plague focus (no. 36) in Russia. *Genome Announc.* **6**, e01440-17 (2018).
79. Zhou, D. et al. Genetics of metabolic variations between *Yersinia pestis* biovars and the proposal of a new biovar, *microtus*. *J. Bacteriol.* **186**, 5147–5152 (2004).
80. Kislichkina, A. A. et al. Six whole-genome assemblies of *Yersinia pestis* subsp. *microtus* bv. *Ulegeica* (phylogroup o.PE5) strains isolated from Mongolian natural plague foci. *Genome Announc.* **6**, e00536-18 (2018).
81. Stamatakis, A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313 (2014).
82. Kembel, S. W. et al. Picante: R tools for integrating phylogenies and ecology. *Bioinformatics* **26**, 1463–1464 (2010).

83. Faith, D. P. Conservation evaluation and phylogenetic diversity. *Biol. Conserv.* **61**, 1–10 (1992).
84. Efron, B. & Tibshirani, R. Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. *Stat. Sci.* **1**, 54–74 (1986).
85. Müllner, D. Fastcluster: fast hierarchical, agglomerative clustering routines for R and Python. *J. Stat. Softw.* **53**, 1–18 (2013).
86. Kumar, S., Stecher, G. & Tamura, K. MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.* **33**, 1870–1874 (2016).
87. Bouckaert, R. et al. BEAST 2.5: an advanced software platform for Bayesian evolutionary analysis. *PLoS Comput. Biol.* **15**, e1006650 (2019).
88. Drummond, A. J., Rambaut, A., Shapiro, B. & Pybus, O. G. Bayesian coalescent inference of past population dynamics from molecular sequences. *Mol. Biol. Evol.* **22**, 1185–1192 (2005).
89. Mühlemann, B. et al. Diverse variola virus (smallpox) strains were widespread in northern Europe in the Viking Age. *Science* **369**, eaaw8977 (2020).
90. Giffin, K. et al. A treponemal genome from an historic plague victim supports a recent emergence of yaws and its presence in 15th century Europe. *Sci. Rep.* **10**, 9499 (2020).
91. Kocher, A. et al. Ten millennia of hepatitis B virus evolution. *Science* **374**, 182–188 (2021).
92. Key, F. M. et al. Emergence of human-adapted *Salmonella enterica* is linked to the Neolithization process. *Nat. Ecol. Evol.* **4**, 324–333 (2020).
93. Sabin, S. et al. A seventeenth-century *Mycobacterium tuberculosis* genome supports a Neolithic emergence of the *Mycobacterium*

- tuberculosis* complex. *Genome Biol.* **21**, 201 (2020).
94. Stadler, T., Kühnert, D., Bonhoeffer, S. & Drummond, A. J. Birth–death skyline plot reveals temporal changes of epidemic spread in HIV and hepatitis C virus (HCV). *Proc. Natl Acad. Sci. USA* **110**, 228–233 (2013).
  95. Darriba, D., Taboada, G. L., Doallo, R. & Posada, D. jModelTest 2: more models, new heuristics and parallel computing. *Nat. Methods* **9**, 772 (2012).
  96. Gorbunova, V., Bozzella, M. J. & Seluanov, A. Rodents for comparative aging studies: from mice to beavers. *Age (Dordr.)* **30**, 111–119 (2008).
  97. Mahmoudi, A. et al. Plague reservoir species throughout the world. *Integr. Zool.* **16**, 820–833 (2021).
  98. Dean, K. R. et al. Human ectoparasites and the spread of plague in Europe during the Second Pandemic. *Proc. Natl Acad. Sci. USA* **115**, 1304–1309 (2018).
  99. Kass, R. E. & Raftery, A. E. Bayes factors. *J. Am. Stat. Assoc.* **90**, 773–795 (1995).
  100. Rambaut, A., Lam, T. T., Max Carvalho, L. & Pybus, O. G. Exploring the temporal structure of heterochronous sequences using TempEst (formerly Path-O-Gen). *Virus Evol.* **2**, vew007 (2016).
  101. Duchene, S. et al. Bayesian evaluation of temporal signal in measurably evolving populations. *Mol. Biol. Evol.* **37**, 3363–3379 (2020).

## Acknowledgements

We thank A. Herbig, G. Neumann and A. Andrades Valtueña and all other members of the Molecular Palaeopathology and Computational Pathogenomics working groups of the Max Planck Institute for Evolutionary Anthropology for helpful discussions throughout the course of this study; C.

Posth and R. Barquera for comments on early versions of the manuscript; T. Hermes for helpful discussions during the review of this study; M. O'Reilly for graphical support; R. I. Tukhbatova, N. Martins, F. Aron, A. Wissgott, R. Radzeviciute and G. Brandt at the Max Planck Institute for the Science of Human History in Jena as well as S. Nagel at the Max Planck Institute for Evolutionary Anthropology in Leipzig for laboratory support; K. Prüfer and S. Clayton for computational assistance. Radiocarbon dating took place at the Curt-Engelhorn-Zentrum Archäometrie in Mannheim, Germany; we thank R. Freidrich for assisting with the interpretation of the radiocarbon dating results. Moreover, we thank V. I. Selezneva from the Peter the Great Museum of Anthropology and Ethnography for initial aid with specimen sampling and N. Smelova, currently at the University of Oslo, for providing essential contextual information during the initial phase of this project. This project received funding from the European Research Council under the European Union's Horizon 2020 research and innovation programme under grant no. 771234 (PALEoRIDER to W.H.). L.M. and L.D. were supported by grant no. AP08856654 from the Ministry of Education and Science of the Republic of Kazakhstan. M.A.S., G.A.G.R., A.K., K.I.B., D.K. and J.K. were also supported by the Max Planck Society.

## Funding

Open access funding provided by Max Planck Society.

## Author information

### Authors and Affiliations

1. Institute for Archaeological Sciences, Eberhard Karls University of Tübingen, Tübingen, Germany

Maria A. Spyrou

2. Department of Archaeogenetics, Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany

Maria A. Spyrou, Lyazzat Musralina, Guido A. Gneccchi  
Ruscone, Arthur Kocher, Kirsten I. Bos, Denise Kühnert, Wolfgang  
Haak & Johannes Krause

3. Department of Archaeogenetics, Max Planck Institute for the Science of Human History, Jena, Germany

Maria A. Spyrou, Lyazzat Musralina, Guido A. Gneccchi  
Ruscone, Arthur Kocher, Kirsten I. Bos, Denise Kühnert, Wolfgang  
Haak & Johannes Krause

4. Laboratory of Population Genetics, Institute of Genetics and Physiology, Almaty, Kazakhstan

Lyazzat Musralina & Leyla Djansugurova

5. Kazakh National University by al-Farabi, Almaty, Kazakhstan

Lyazzat Musralina

6. Transmission, Infection, Diversification & Evolution Group, Max Planck Institute for the Science of Human History, Jena, Germany

Arthur Kocher & Denise Kühnert

7. Department of Civilisations and Forms of Knowledge, University of Pisa, Pisa, Italy

Pier-Giorgio Borbone

8. Department of Physical Anthropology, Kunstkamera, Peter the Great Museum of Anthropology and Ethnography, Russian Academy of Sciences, St Petersburg, Russian Federation

Valeri I. Khartanovich

9. Research Institute and Museum of Anthropology, Lomonosov Moscow State University, Moscow, Russian Federation

Alexandra Buzhilova

10. European Virus Bioinformatics Center (EVBC), Jena, Germany

Denise Kühnert

11. Division of History, Heritage and Politics, University of Stirling,  
Stirling, UK

Philip Slavin

## Contributions

M.A.S., P.S. and J.K. conceived and led the investigation. M.A.S., K.I.B., P.S. and J.K. designed the study. M.A.S. and L.M. performed the laboratory work. M.A.S. and A.K. performed the bacterial genomic data analysis. M.A.S., A.K. and D.K. performed the molecular dating analysis. G.A.G.R. performed the human population genetic analysis. P.-G.B. and P.S. assembled, analysed and translated the historical, archaeological and epigraphic context information. A.B. and V.I.K. provided access to the archaeological material and contextual information. M.A.S., A.K., L.D., K.I.B., D.K., W.H., P.S. and J.K. aided in interpreting the results. M.A.S. and P.S. wrote the paper with contributions from all co-authors.

## Corresponding authors

Correspondence to [Maria A. Spyrou](#), [Philip Slavin](#) or [Johannes Krause](#).

## Ethics declarations

## Competing interests

The authors declare no competing interests.

## Peer review

## Peer review information

*Nature* thanks Ludovic Orlando, Mark Achtman and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. [Peer reviewer reports](#) are available.

## Additional information

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

## Extended data figures and tables

### [Extended Data Fig. 1 Available tombstone pictures from Kara-Djigach.](#)

**a–c**, Available tombstone pictures from individuals investigated as part of this study. For a translation of the tombstone inscriptions, see individual descriptions within Supplementary Information 2. Tombstone dates are as follows: Grave 9 (1338-9 CE), Grave 19/20 (1338-9 CE), Grave 6 (year not inscribed). Picture credits to P.-G. Borbone. **d+e**, Tombstones identified in Kara-Djigach containing pestilence-stating inscriptions, dating to the years 1338 and 1339 CE. These tombstones do not correspond to individuals analysed within our aDNA dataset. The original tombstone on panel d (without traced inscription) is shown in Fig. 1. Complete translations are available within Supplementary Information 2.

### [Extended Data Fig. 2 Ancient DNA damage substitution frequencies for all \*Y. pestis\* captured libraries.](#)

C-to-T substitution frequencies characteristic of post-mortem deamination of ancient DNA are shown for the 5' ends of sequenced reads aligned against the CO92 *Y. pestis* reference genome (NC\_003143.1).

### **Extended Data Fig. 3 Evaluation of BSK007 after whole-genome *Y. pestis* capture.**

**a**, Post-capture coverage distribution of BSK007 across the *Y. pestis* CO92 chromosome. Mean coverage was estimated across the genome in 4,000 bp windows. The dotted gray line indicates the mean coverage across the entire genome (0.125-fold). **b**, Krona plots showing the taxonomic classification of BSK007 reads mapping against all *Y. pestis* CO92 elements (chromosome NC\_003143.1, pMT1 NC\_003134.1, pPCP1 NC\_003132.1 and pCD1 NC\_003131.1). Numbers in brackets next to element designations correspond to the number of assigned reads in MALT. The colours of Krona sectors represent different taxonomic levels and their completeness is proportional to the relative abundance of summarised reads at each corresponding taxonomic node. The shown percentages indicate the species-level (outermost circle) proportion of reads aligned to taxa other than *Y. pestis* and *Y. pseudotuberculosis*.

### **Extended Data Fig. 4 Read length and ancient DNA damage distribution of contaminant SNP regions in the combined BSK001/003 dataset prior to metagenomic filtering.**

Overlaid length distributions of reads mapping against the CO92 *Y. pestis* reference genome, calculated for the entire dataset (gray) as well as for 117 regions surrounding putatively contaminant SNPs (blue). Regions were extracted within a 150 bp window surrounding each putatively contaminant SNP. Dotted lines represent average fragment lengths for the entire dataset and for the 117 putatively contaminant SNP regions in gray and blue, respectively. Reads comprising contaminant SNP regions show a distinct length distribution compared to the one observed across the entire BSK001/003 genome, with a marked shift towards longer read lengths. The 76 bp fragment length peak represents the uppermost possible read length of single-end sequenced reads, which comprised the majority of data within the present dataset. Ancient DNA damage patterns were compared between the entire dataset (upper panel) and the putatively contaminant SNP regions (lower panel), showing a near 3-fold reduction in the latter as estimated for the terminal 5' base.

## Extended Data Fig. 5 Phylogenetic comparisons between BSK001/003 against ancient and modern *Y. pestis* diversity.

Full length maximum likelihood phylogenetic tree using 1,000 bootstrap iterations for estimating node support and visualised using FigTree v1.4.4. The tree is constructed with 203 modern and 48 historical *Y. pestis* genomes, and is based on 2,960 SNPs (96% partial deletion). Scale denotes the number of substitutions per genomic site.

## Extended Data Fig. 6 Evaluation of phylogenetically diagnostic SNPs across 14<sup>th</sup> century *Y. pestis* genomes.

The estimated variant calls were retrieved from a SNP table comprising 203 modern and 48 historical *Y. pestis* genomes (full dataset contains 3,533 SNPs). Error bars indicate uncertainty due to the presence of missing data (Ns) within the variant calls of the respective genomes.

## Extended Data Fig. 7 Comparison of phylogenetic diversity measures computed for the complete modern *Y. pestis* phylogeny against the Branch 1-4 subclade.

The left panels show maximum likelihood substitution trees considering only extant *Y. pestis* genomes, annotated based on the compared clades and their corresponding Faith's phylogenetic diversity index (FPD).

Phylogenetic branches considered for the Branch 1-4 FPD computation are shown in green (full dataset) and red (subsampled dataset). The right panels show violin plots indicating the distribution and mean of pairwise phylogenetic distances based on maximum likelihood trees (MPD). Median estimates and 95% percentile intervals were derived from the resampled bootstrap trees (1,000 bootstrap iterations). Points within violin plots indicate the mean estimated phylogenetic distance for all datasets.

## Extended Data Fig. 8 Phylogenetic relationships of 0.ANT lineages.

Maximum likelihood phylogenetic tree based on 2,441 genome-wide variant positions (all SNPs). The tree was constructed to indicate the genetic relationships between all previously published extant O.ANT genomes and BSK001/003. Scale denotes the number of substitutions per genomic site. Node support was determined by 1,000 bootstrap iterations.

## Supplementary information

### Supplementary Information

Supplementary Sections 1–4 and References.

### Reporting Summary

### Peer Review File

### Supplementary Tables 1–22.

## 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

Spyrou, M.A., Musralina, L., Gnechi Ruscone, G.A. *et al.* The source of the Black Death in fourteenth-century central Eurasia. *Nature* **606**, 718–724 (2022). <https://doi.org/10.1038/s41586-022-04800-3>

- Received: 19 July 2021
- Accepted: 25 April 2022
- Published: 15 June 2022
- Issue Date: 23 June 2022
- DOI: <https://doi.org/10.1038/s41586-022-04800-3>

## Share this article

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

[Get shareable link](#)

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

[Copy to clipboard](#)

Provided by the Springer Nature SharedIt content-sharing initiative

## Further reading

- [Ancient DNA traces origin of Black Death](#)

- Ewen Callaway

*Nature* (2022)

## [How the Black Death got its start](#)

- Nick Petrić Howe
- Shamini Bundell

## Nature Nature Podcast 15 Jun 2022

---

This article was downloaded by **calibre** from <https://www.nature.com/articles/s41586-022-04800-3>

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

- Article
- [Published: 12 May 2022](#)

# The role of NSP6 in the biogenesis of the SARS-CoV-2 replication organelle

- [Simona Ricciardi](#),
- [Andrea Maria Guarino](#),
- [Laura Giaquinto](#),
- [Elena V. Polishchuk](#),
- [Michele Santoro](#),
- [Giuseppe Di Tullio](#),
- [Cathal Wilson](#),
- [Francesco Panariello](#),
- [Vinicius C. Soares](#),
- [Suelen S. G. Dias](#),
- [Julia C. Santos](#),
- [Thiago M. L. Souza](#),
- [Giovanna Fusco](#),
- [Maurizio Viscardi](#),
- [Sergio Brandi](#),
- [Patrícia T. Bozza](#),
- [Roman S. Polishchuk](#) ,
- [Rossella Venditti](#) &
- [Maria Antonietta De Matteis](#)

[Nature](#) volume **606**, pages 761–768 (2022)

- 10k Accesses
- 72 Altmetric

- [Metrics details](#)

## Abstract

SARS-CoV-2, like other coronaviruses, builds a membrane-bound replication organelle to enable RNA replication<sup>1</sup>. The SARS-CoV-2 replication organelle is composed of double-membrane vesicles (DMVs) that are tethered to the endoplasmic reticulum (ER) by thin membrane connectors<sup>2</sup>, but the viral proteins and the host factors involved remain unknown. Here we identify the viral non-structural proteins (NSPs) that generate the SARS-CoV-2 replication organelle. NSP3 and NSP4 generate the DMVs, whereas NSP6, through oligomerization and an amphipathic helix, zippers ER membranes and establishes the connectors. The NSP6( $\Delta$ SGF) mutant, which arose independently in the Alpha, Beta, Gamma, Eta, Iota and Lambda variants of SARS-CoV-2, behaves as a gain-of-function mutant with a higher ER-zippering activity. We identified three main roles for NSP6: first, to act as a filter in communication between the replication organelle and the ER, by allowing lipid flow but restricting the access of ER luminal proteins to the DMVs; second, to position and organize DMV clusters; and third, to mediate contact with lipid droplets (LDs) through the LD-tethering complex DFCP1–RAB18. NSP6 thus acts as an organizer of DMV clusters and can provide a selective means of refurbishing them with LD-derived lipids. Notably, both properly formed NSP6 connectors and LDs are required for the replication of SARS-CoV-2. Our findings provide insight into the biological activity of NSP6 of SARS-CoV-2 and of other coronaviruses, and have the potential to fuel the search for broad antiviral agents.

## Main

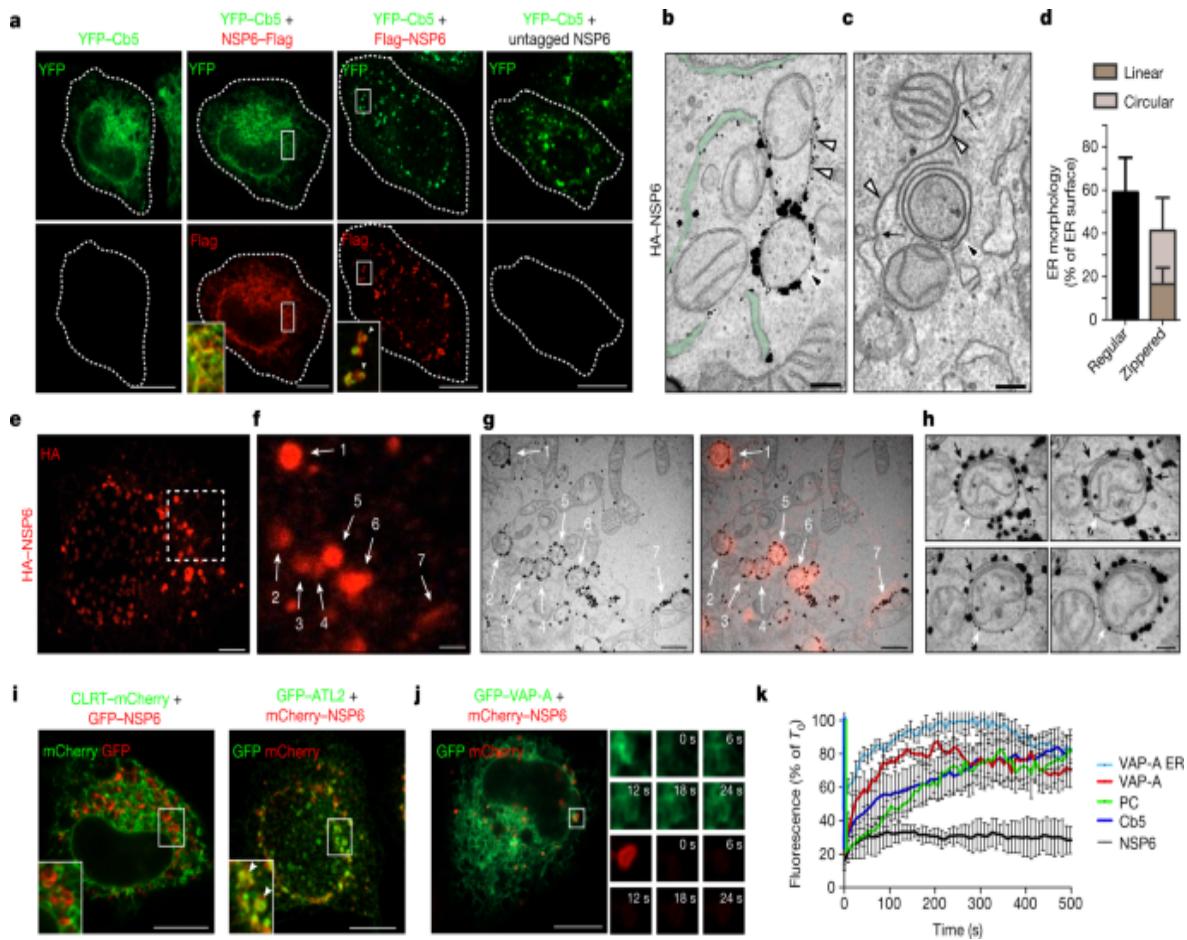
SARS-CoV-2 extensively rearranges host cellular membranes into replication organelles that provide a microenvironment conducive to RNA synthesis and protection from host sensor and defence systems<sup>1,2</sup>. The 16 viral NSPs that are released from polyproteins pp1a and pp1ab by 2 viral proteases include 13 cytosolic proteins, which are involved in RNA replication, and 3 transmembrane proteins, NSP3, NSP4 and NSP6. Studies

on other coronaviruses suggest that NSP3 and NSP4, with a hitherto undefined contribution from NSP6, are responsible for generating the replication organelles<sup>3,4,5,6</sup>. Despite considerable advances in our understanding of the ultrastructure of the SARS-CoV-2 replication organelle<sup>2,7,8</sup>, mechanistic insights into its biogenesis have so far been limited. In particular, there is at present—to our knowledge—no information on the role of NSP6 in this process. Of note, six SARS-CoV-2 ‘variants of concern’ (VOCs) (Alpha (B.1.1.7), Beta (B.1.351), Gamma (P.1), Eta (B.1.525), Iota (B.1.526)<sup>9</sup> and Lambda (C.37)<sup>10</sup>) share a three-amino-acid deletion in NSP6 (NSP6( $\Delta$ SGF)), in addition to the more noted mutations in the spike protein; this finding adds further impetus to the need to examine the role of NSP6 in the biogenesis of replication organelles and in the replication of SARS-CoV-2.

## NSP6 induces ER zippering

We tagged SARS-CoV-2 NSP6 at either the N or the C terminus. C-terminally tagged NSP6 showed a diffuse distribution in the ER (Fig. 1a and Extended Data Fig. 1a), as reported for NSP6 from other coronaviruses<sup>3,11</sup>. Conversely, N-terminally tagged NSP6, expressed at a comparable level, and untagged NSP6 elicited the formation of roundish structures (Fig. 1a and Extended Data Fig. 1a–c). These structures, which we refer to as the NSP6 compartment, did not colocalize with endosomal, lysosomal or autophagosomal markers (Extended Data Fig. 1d), but colocalized with the ER-reporter protein Cb5 (the C-terminal tail of cytochrome *b*5)<sup>12</sup> (Fig. 1a and Extended Data Fig. 1a). This appears to be a general feature of coronavirus NSP6, as N- but not C-terminally tagged avian infectious bronchitis virus (IBV) NSP6 also formed roundish structures that colocalized with Cb5 (Extended Data Fig. 1e).

**Fig. 1: NSP6 induces ER zippering.**



**a**, HeLa cells expressing YFP–Cb5 alone, co-expressing C- or N-terminally Flag-tagged NSP6, or co-expressing untagged NSP6. Insets, enlarged merged images of boxed areas; arrowheads, NSP6 compartments; dashed lines, cell boundaries. **b,c**, IEM (anti-haemagglutinin (HA) immunolabelling) (**b**) and EM (**c**) of HeLa cells expressing HA–NSP6. White arrowheads, linear zippered ER membranes; black arrowheads, circular zippered ER membranes; black arrows, continuity between zippered and regular ER membranes. The regular ER is shown in green. The average size of circular NSP6-positive ER structures is  $623 \pm 231$  nm. **d**, Morphometric analysis of NSP6-expressing cells (percentage of the ER surface that is associated with regular cisternae or zippered domains). Mean  $\pm$  s.d.,  $N = 3$ ,  $n = 60$  (in all figure legends,  $N$  indicates the number of experiments and  $n$  the number of total measurements or observations). **e–h**, Immuno-CLEM analysis of the NSP6 compartments. **e**, Fluoromicrograph of HA–NSP6. **f,g**, Enlargement with NSP6-labelled structures 1–7 that were identified on EM serial sections (**g**, left), and correspondence of NSP6

fluorescent spots with NSP6 circular and linear zippered ER membranes (**g**, right). **h**, Serial sections of structure 6 in **g**. Black arrows, NSP6-positive linear zippered membrane connections with ER cisternae. White arrows, NSP6-positive circular zippered structures. **i**, HeLa cells co-expressing NSP6 with CLRT or ATL2 as indicated. Insets, enlarged merged images of boxed areas; arrowheads indicate co-localization. **j,k**, FRAP analysis of GFP–VAP-A and mCherry–NSP6 co-expressing cells. **j**, NSP6 compartments (boxed) were photobleached and the fluorescence recovery was monitored. The small panels are representative frames showing the time in seconds after the bleach (see Supplementary Video 2). **k**, Quantitative FRAP analysis of the reporters in **j** and in Extended Data Fig. 2e,f. VAP-A ER, FRAP of VAP-A in areas of regular ER; PC, BODIPY C<sub>12</sub>-HPC. Fluorescence intensity is expressed as a percentage of the value measured at time 0 ( $T_0$ ), which is normalized to 100%. Mean  $\pm$  s.d.,  $N = 3$ ,  $n = 45$ . Scale bars, 10  $\mu\text{m}$  (**a,i,j**); 120 nm (**b,c**); 4  $\mu\text{m}$  (**e**); 1.1  $\mu\text{m}$  (**f,g**); 250 nm (**h**).

### Source data

Immunoelectron microscopy (IEM) showed that NSP6 was highly concentrated on ER cisternae that had tightly juxtaposed delimiting membranes, leaving a barely visible lumen (Fig. 1b and Extended Data Fig. 1f,g). These structures—which we refer to as zippered ER, by analogy with those reported for other viruses<sup>4,6</sup>—were highly reminiscent of the 'ER connectors' that are observed in SARS-CoV-2-infected cells<sup>2</sup>. The zippered ER structures were either linear or, more often, circular structures that encapsulated the neighbouring cytoplasm (Fig. 1c,d). Clear connections between these zippered ER structures and the regular ER were visible by electron microscopy (EM) and electron tomography (Fig. 1c, Extended Data Fig. 1h–j and Supplementary Video 1). Correlative light and electron microscopy (CLEM) showed that the roundish or elongated NSP6 spots that were visualized by immunofluorescence corresponded, respectively, to the circular or linear zippered ER profiles that were observed by EM (Fig. 1e–g), whose connection to the regular ER can be traced (Fig. 1h). The NSP6-containing structures were not freely accessible to ER luminal proteins (such as calreticulin and the ER reporter GFP-KDEL) or to ER membrane proteins with bulky luminal domains (such as ERGIC53 and ATF6) but were accessible to ER membrane proteins such as VAP-A, atlastin-2 and the

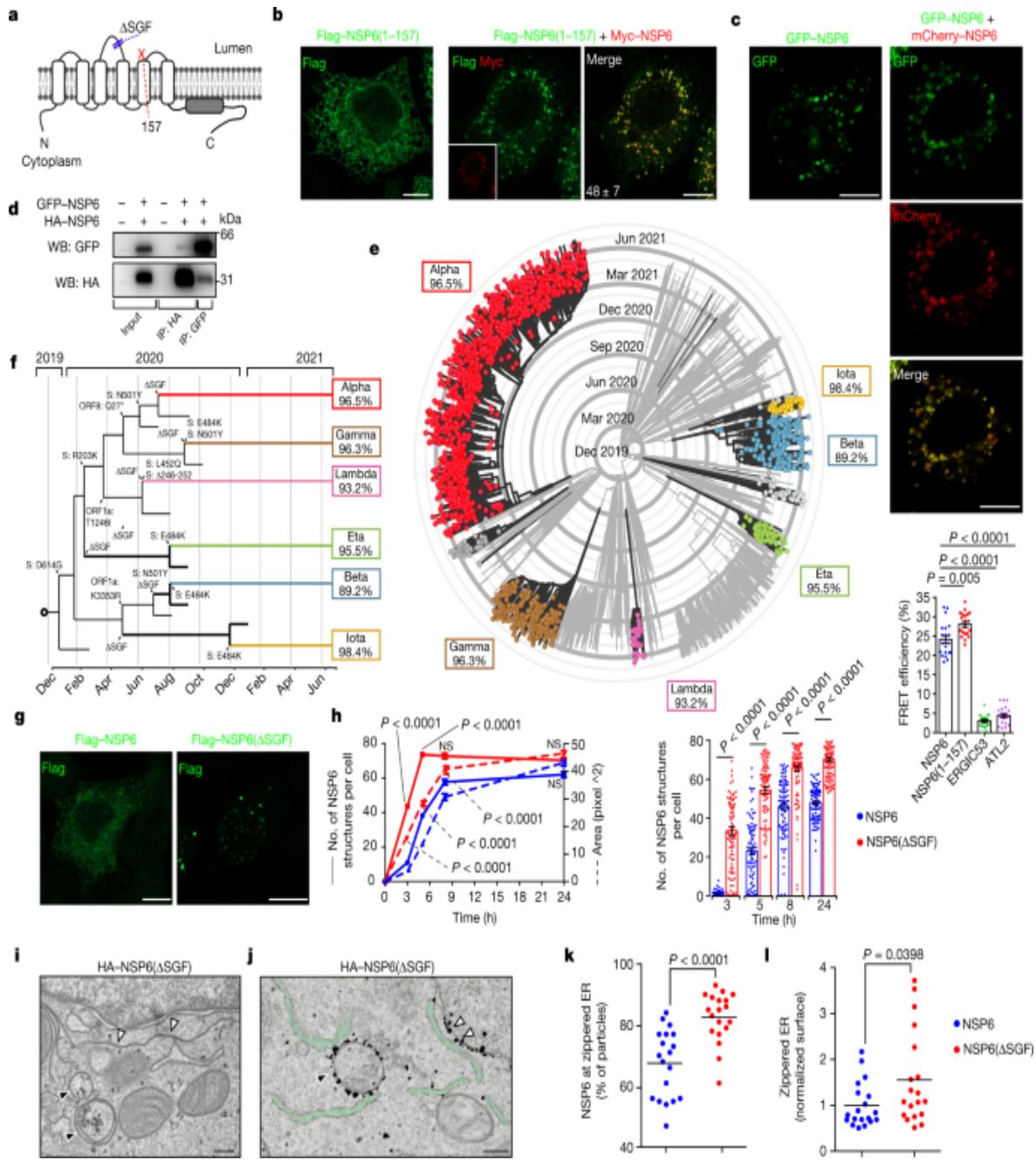
KDEL receptor that possess no or very small luminal tracts (Fig. 1*i–k* and Extended Data Fig. 2*a–d*). We validated the continuity between the NSP6 compartment and the ER using fluorescence recovery after bleaching (FRAP) assays (Fig. 1*j,k*, Extended Data Fig. 2*e* and Supplementary Videos 2 and 3). After bleaching, both VAP-A and Cb5 re-entered the NSP6 compartment, although with slower kinetics compared with the regular ER. The NSP6 compartment was accessible to phospholipids, such as BODIPY C<sub>12</sub>-HPC, whose fluorescence also recovered after bleaching (Fig. 1*k*, Extended Data Fig. 2*f* and Supplementary Video 4). NSP6 itself, however, showed limited FRAP (Fig. 1*j,k* and Extended Data Fig. 2*e,f*), probably because it is engaged in stable protein complexes.

Our results indicate that NSP6 drives the formation of a zippered double-membrane compartment that maintains continuity with the ER but largely excludes ER luminal proteins.

## NSP6 homodimers zipper ER membranes

The structure of NSP6 has not been solved and different topologies have been predicted. The N and C termini of NSP6 must face the cytosol because they are processed by the cytosolic NSP5 protease. Indeed, N- or C-terminally tagged NSP6 was readily detectable by antibodies after plasma membrane permeabilization (Extended Data Fig. 3*a*). On the basis of these data, topology predictions using the CCTOP server<sup>13</sup> and biochemical analyses of other coronaviruses<sup>14,15</sup>, we assigned six transmembrane domains to NSP6 and envisaged that the seventh predicted transmembrane domain, which is an amphipathic helix<sup>16</sup>, does not cross, but remains associated with, the membrane (Fig. 2*a* and Extended Data Fig. 3*b*).

**Fig. 2: ER zippering requires NSP6 homodimerization and is more efficient with NSP6(ΔSGF).**



**a**, Predicted secondary structure of NSP6. The ΔSGF deletion and truncation site (at residue 157) are indicated. **b**, HeLa cells expressing Myc-NSP6 (inset) and/or Flag-NSP6(1–157). The fraction of NSP6(1–157) that is associated with NSP6 structures is indicated. Mean  $\pm$  s.d.,  $N=3$ ,  $n=74$ . **c**, Cells expressing GFP-NSP6 alone or together with mCherry-NSP6. Bottom graph, FRET measurements in cells co-expressing mCherry-NSP6 with the indicated GFP-tagged protein. Mean  $\pm$  s.d.,  $N=3$ ,  $n=20$ . **d**, Immunoprecipitation (IP) and western blot (WB) from GFP-NSP6 and HA-

NSP6 co-expressing cells, representative of four independent experiments. **e**, Radial layout of a phylogenetic tree of 3,508 SARS-CoV-2 genomes. VOCs are indicated and the percentage of each genome containing ΔSGF is reported. Black branches highlight the appearance of the deletion. **f**, Mutations involved in the branching and specificity of each VOC. Arrows, appearance of the ΔSGF and mutations in the spike (S) protein. **g,h**, Time course analysis of stably expressing Flag–NSP6 or Flag–NSP6(ΔSGF) cells induced with doxycycline. **g**, Fluoromicrographs at 3 h. **h**, Quantification of the structures shown in **g** and Extended Data Fig. [5a](#) ( $N=3$ ,  $n=90$ ). Left, number and areas of NSP6-positive structures. One pixel corresponds to  $0.0825\text{ }\mu\text{m}$ . Right, NSP6 in structures as a percentage of the total NSP6 in the cell (mean  $\pm$  s.e.m.). **i,j**, EM (**i**) and IEM (**j**) (anti-HA immunolabelling) of HA–NSP6(ΔSGF)-expressing HeLa cells. White arrowheads, linear zippered ER structures; black arrowheads, circular zippered ER structures. The regular ER is shown in green. **k**, Morphometric analysis of IEM images. Quantification of gold particles at zippered ER (percentage of the total ER-associated-particles). **l**, The surface area of zippered ER normalized for the total number of gold particles. For **k,l**,  $N=3$ ,  $n=19$ . Scale bars,  $10\text{ }\mu\text{m}$  (**b,c,g**);  $250\text{ nm}$  (**i,j**). Two-tailed Mann–Whitney test (**c**); unpaired two-tailed *t*-test (**k,l**); one-way ANOVA with Tukey’s post-hoc test (**h**). NS, not significant.

## Source data

Truncating the C-terminal part of NSP6 including the amphipathic helix (NSP6(1–157); Fig. [2a](#)) or introducing two mutations that abrogate its amphiphilic properties<sup>16</sup> (NSP6(F220Q/T222W); Extended Data Fig. [3b](#)) caused NSP6 to distribute diffusely in the ER (Fig. [2b](#) and Extended Data Fig. [3c,d](#)). However, although necessary, the amphipathic helix is not sufficient to induce ER remodelling as the C-terminal domain—which includes part of the amphipathic helix (NSP6-C80; see below)—was unable to induce it. We found that homodimerization of NSP6 is also required. Fluorescence resonance energy transfer (FRET) measurements and the co-immunoprecipitation of GFP–NSP6 co-expressed with mCherry–NSP6 or HA–NSP6 indicated that NSP6 undergoes homodimerization (Fig. [2c,d](#) and Extended Data Fig. [3e,f](#)). Dimerization involves amino acids 1–157, as NSP6(1–157) was massively recruited and retained in the NSP6 compartment when co-expressed with the full length NSP6 (Fig. [2b](#)).

Indeed, both FRET and co-immunoprecipitation experiments (Fig. 2c and Extended Data Fig. 3g,h) confirmed that NSP6(1–157) and NSP6 interact, which indicates that NSP6(1–157) maintains the homodimerization interface(s).

Together, these data indicate that both the C-terminal amphipathic helix and NSP6 homodimerization (through amino acids 1–157) are required to generate the NSP6 compartment.

K22, a small molecule that interferes with the replication of several coronaviruses with different potencies, has been hypothesized to target NSP6, as K22-resistant strains of human coronavirus 229E (HCoV-229E) have mutations in NSP6 (ref. 17). We found that K22 (at 40 µM) reduced the number of regular NSP6 structures and NSP6 retention in these structures (Extended Data Fig. 4a,b). In addition, 37% of K22-treated cells exhibited elongated perinuclear NSP6 structures (Extended Data Fig. 4a,c). ImmunocLEM (Extended Data Fig. 4d–f) and EM (Extended Data Fig. 4g–j) showed that these structures corresponded to extensive zippered areas of the nuclear envelope. Thus, impaired formation of the NSP6 compartment induced by K22 may be due in part to a shift in NSP6 zippering activity towards the nuclear envelope, which is apparently an unfavourable site for replication organelle formation (<https://www.ebi.ac.uk/empiar/EMPIAR-10490/>).

## NSP6( $\Delta$ SGF) has a higher ER-zippering activity

Six SARS-CoV-2 VOCs (Alpha, Beta, Gamma, Eta, Iota and Lambda) have a three-amino-acid deletion (SGF, positions 106–108) in the predicted second and longest NSP6 luminal loop. Phylogenetic analysis of SARS-CoV-2 using Nextstrain<sup>18</sup> showed that the deletion emerged independently in these lineages (Fig. 2e,f), suggesting that it conveys a selective advantage.

We found that, compared with the Wuhan-Hu-1 NSP6 (from here on, the reference NSP6), NSP6( $\Delta$ SGF) is more proficient in inducing the NSP6 compartment: the kinetics of formation were faster (Fig. 2g,h, Extended Data Fig. 5a); the NSP6( $\Delta$ SGF) compartments were more numerous and larger (Fig. 2h); and NSP6( $\Delta$ SGF) was more enriched in these compartments

(Fig. 2*h*). These differences were not due to different protein levels or half-lives (Extended Data Fig. 5*b,c*), but to the higher propensity of NSP6( $\Delta$ SGF) to homo-oligomerize, as indicated by its higher resistance to detergent extraction (Extended Data Fig. 5*d*), more efficient co-immunoprecipitation (Extended Data Fig. 5*e*) and lower mobility (in fluorescence loss in photobleaching (FLIP) and FRAP experiments) compared to the reference NSP6 (Extended Data Fig. 5*f–h*). NSP6( $\Delta$ SGF) was slightly less sensitive than the reference NSP6 to K22 (Extended Data Fig. 5*g–i*). Finally, EM, IEM and CLEM showed that NSP6( $\Delta$ SGF) promoted the formation of both linear and circular zippered membrane compartments (Fig. 2*j,l* and Extended Data Fig. 5*j–l*), like NSP6, but that it was more highly associated with zippered membrane domains and was depleted from the regular ER (Fig. 2*j,k*). This was paralleled by an increase in the ER surface area occupied by zippered domains (Fig. 2*l*).

The higher ER-zippering activity of NSP6( $\Delta$ SGF) was also evident comparing the putative precursor of NSP6; that is, NSP6–NSP7 and NSP6( $\Delta$ SGF)–NSP7. During viral infection, NSP6 is generated by polyprotein cleavage by NSP5 (ref. 1). Consistent with NSP6 forming the NSP6 compartment only if its C terminus is ‘free’, NSP6–NSP7 showed a diffuse ER distribution (Extended Data Fig. 5*m,n*) and also a partial Golgi localization, suggesting that the precursor might visit the Golgi before the cleavage unleashes its ER-zippering activity. By contrast, NSP6( $\Delta$ SGF)–NSP7 was mainly retained in the ER and was able to form small roundish structures even before cleavage (Extended Data Fig. 5*m,n*).

## NSP6 connects DMVs with the ER

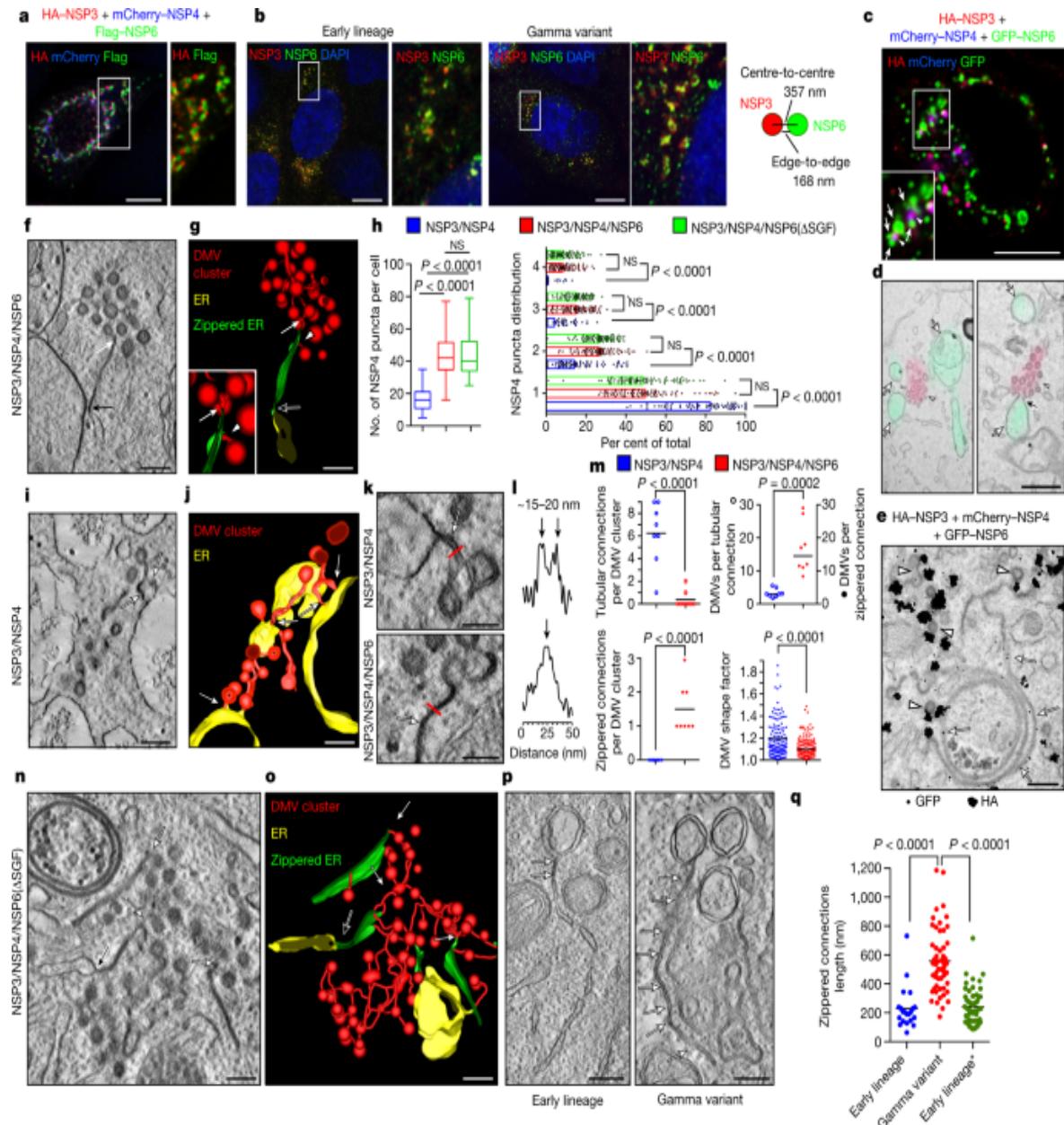
Given the similarity of NSP6-induced zippered ER to ‘ER connectors’ between the ER and DMVs in SARS-CoV-2-infected cells<sup>2</sup>, we investigated the relationship between NSP6 and the DMVs. When expressed alone, NSP3 and NSP4 exhibited a diffuse distribution in the ER (Extended Data Fig. 6*a*), but when co-expressed (hereafter referred to as NSP3/NSP4), and in agreement with recent reports<sup>19,20</sup>, they fully colocalized in punctate structures (Extended Data Fig. 6*b,c*). At the EM level, these corresponded to clusters of vesicles that had a diameter of 50–100 nm and were surrounded

by two membranes (that is, DMVs), with a visible intermembrane space (Extended Data Fig. [6d,e](#)).

Thus, NSP3/NSP4 and NSP6 are individually able to reproduce the two main features of the SARS-CoV-2 replication organelle, DMVs and the connectors<sup>2</sup>, respectively.

The combined expression of all three membrane NSPs (NSP3/NSP4/NSP6; Extended Data Fig. [6f,g](#)) revealed NSP3/NSP4-positive puncta in close proximity to but not overlapping the NSP6 compartment (Fig. [3a](#)). Notably, a similar segregation of NSP6- and NSP3-positive domains was also detectable in Calu-3 cells infected with an early lineage or the Gamma variant of SARS-CoV-2 (Fig. [3b](#)). CLEM revealed that the NSP3/NSP4 puncta corresponded to clusters of DMVs, whereas the NSP6 structures corresponded to tracts of zippered ER that remained distinct from but were often close and connected to the DMVs (Fig. [3c,d](#)). IEM showed that groups of NSP3/NSP4-positive DMVs associated with NSP6-positive zippered ER membranes (Fig. [3e](#)). Tomographic analysis of NSP3/NSP4/NSP6-expressing cells revealed that the DMVs were organized in 'grape-like' clusters, sometimes with reciprocal connections (Fig. [3f,g](#)). Long tracts of zippered ER formed connections between the DMV clusters and the ER proper (Fig. [3f,g](#) and Supplementary Videos [5–7](#)), similar to those observed in SARS-CoV-2-infected cells<sup>2</sup>. Thus, we inferred that NSP6 forms the zippered connectors that guarantee full membrane—but restricted luminal—continuity with the ER.

**Fig. 3: NSP6-zippered membranes connect NSP3/NSP4 DMVs to the ER.**



**a**, Calu-3 cells expressing HA-NSP3 + Flag-NSP6 + mCherry-NSP4. The enlargement of the boxed area shows HA and Flag immunolabelling. **b**, Calu-3 cells infected with early lineage and Gamma variant SARS-CoV-2. Values represent mean NSP3- and NSP6-structure distances in nm.  $N=10$  cells,  $n=2,233$  structures. **c,d**, CLEM. Fluoromicrograph (c) and EM serial sections (d) of an HA-NSP3 + mCherry-NSP4 + GFP-NSP6-expressing cell. Arrowheads, NSP3/NSP4 co-localization; arrows, NSP6 compartments; black arrow, NSP6 compartment connection with NSP3/NSP4 DMVs. **e**, IEM showing NSP3 (anti-HA, white arrowheads) and NSP6 (anti-GFP, black arrowheads) in Calu-3 cells expressing HA-NSP3 + mCherry-NSP4 + GFP-NSP6.

white arrows). **f,g**, Tomogram (**f**) and three-dimensional (3D) reconstruction (**g**) showing connections of zippered ER to DMVs (white arrow and arrowhead) and to the regular ER (black arrow) (Supplementary Videos 5 and 6). **h**, Number and distribution of NSP4 puncta in cells expressing the indicated NSPs. Mean  $\pm$  s.d.,  $N = 3$ ,  $n = 30$ . Box plots represent 25th to 75th percentiles of the data (centre line, median; whiskers, minima and maxima). **i,j**, Tomograms (**i**) and 3D reconstruction (**j**) showing multiple short DMV–ER tubular connections (white arrows) in an NSP3/NSP4-expressing cell (Supplementary Videos 8 and 9). **k**, Tomograms from NSP3/NSP4 and NSP3/NSP4/NSP6-expressing cells showing DMV–ER connections (arrows). **l**, Intensity profiles along the red lines in **k**. **m**, Morphometry of NSP3/NSP4- or NSP3/NSP4/NSP6-expressing cells. **n,o**, Tomogram (**n**) and 3D reconstruction (**o**) of a NSP3/NSP4/NSP6( $\Delta$ SGF)-expressing cell, showing numerous zippered ER domains connected to DMVs (white arrows) and to regular ER (black arrows) (Supplementary Videos 11 and 12). **p**, Tomograms showing zippered ER connections (arrows) to DMVs in Calu-3 cells infected with early lineage and Gamma variant SARS-CoV-2 (Supplementary Videos 13 and 14). **q**, Length of zippered ER connected to DMVs in infected cells ( $N = 10$ ,  $n \geq 20$ ), and from early lineage\* (Bavpat1/2020) data in EMPIAR-10490 (29 tomograms). Scale bars, 10  $\mu\text{m}$  (**a,b**); 10  $\mu\text{m}$  (**c**); 470 nm (**d**); 200 nm (**e,p**); 160 nm (**f,g,i,j,n,o**); 100 nm (**k**). One-way ANOVA with Tukey's (**h**, left) or Emmeans (**h**, right) post-hoc test; unpaired two-tailed *t*-test (**m,q**). NS, not significant.

### Source data

We then asked how NSP3/NSP4-induced DMVs might be affected by NSP6. The NSP3/NSP4 puncta were more numerous and more homogeneously distributed throughout the cytoplasm in NSP3/NSP4/NSP6-expressing (both reference and NSP6( $\Delta$ SGF)) than in NSP3/NSP4-expressing cells (Fig. 3h), suggesting that NSP6 might provide a cue for the positioning and organization of DMVs (Fig. 3a,c).

Electron tomography revealed that in the absence of NSP6, DMV connections with the ER were short and tubular with a clearly detectable lumen (Fig. 3i–m, Extended Data Fig. 6h and Supplementary Videos 8–10). By contrast, in the presence of NSP6, DMV clusters were connected with the ER through much longer sheet-like zippered domains (Fig. 3f,g,k–m,

Extended Data Fig. 6*h* and Supplementary Videos 5–7). The number of DMVs per connection was also different: an average of around 3 DMVs per tubular connection without NSP6 and of around 15 DMVs per zippered connection with NSP6 (Fig. 3*m*). In addition, the shape of the DMVs was more regular (Fig. 3*m*), their size was more uniform (Extended Data Fig. 6*i–k*) and their packing inside each cluster was denser in the presence of NSP6 (Fig. 3*f,g,i,j* and Extended Data Fig. 6*l–n*).

These data indicate that the co-expressed NSP3/NSP4/NSP6 reproduce replication-organelle-like structures (ROLS) and that NSP6 organizes DMV clusters.

We then assessed whether conditions that negatively or positively affect the ER-zippering activity of NSP6—that is, K22 treatment or SGF deletion, respectively—have an effect on the ROLS. Although K22 had no effect on the number and distribution of NSP3/NSP4 puncta in cells expressing only NSP3 and NSP4, it blunted the ability of co-expressed NSP6 to increase the number of NSP3/NSP4 puncta (Extended Data Fig. 7*a,b*). EM revealed that DMV clusters in these cells contained a significantly lower number of vesicles (Extended Data Fig. 7*c–f*) with a less regular shape (Extended Data Fig. 7*e,f*) that lost zippered connections and acquired more tubular connections to the ER (Extended Data Fig. 7*d–f*). Thus, treatment with K22 counteracted the ability of NSP6 to form zippered connections and to promote the homogeneous growth of DMVs. Corroborating these results, we found that K22, at the (relatively high) concentrations that interfere with the biogenesis of ROLS (40 μM), but not at lower ones<sup>21</sup>, inhibited the replication of SARS-CoV-2 (Extended Data Fig. 7*g–i*).

As for the SGF deletion, we found that NSP6(ΔSGF) also enhances and organizes the formation of NSP3/NSP4 puncta (Fig. 3*h*), but that each DMV cluster contains a higher number of DMVs that are more homogeneous in terms of size, as well as exhibiting a more developed system of zippered connections compared to the reference NSP6 (Fig. 3*n,o*, Extended Data Fig. 8*a–e* and Supplementary Videos 11 and 12).

Finally, we analysed the zippered connectors in Calu-3 cells infected with an early lineage or Gamma variant SARS-CoV-2 that contains the SGF deletion in NSP6, and found that the Gamma strain has a much more extensive

zippered connector system joining the DMVs with each other and with the ER (Fig. 3p–q, Extended Data Fig. 8f–h and Supplementary Videos 13 and 14). One might speculate that the higher zippering activity of NSP6(ΔSGF) has a role in establishing a more functional and better shielded replication organelle, providing one of the multiple mechanisms that contribute to the reported differences in replication dynamics and immune evasion of NSP6(ΔSGF)-bearing VOCs<sup>22,23</sup>.

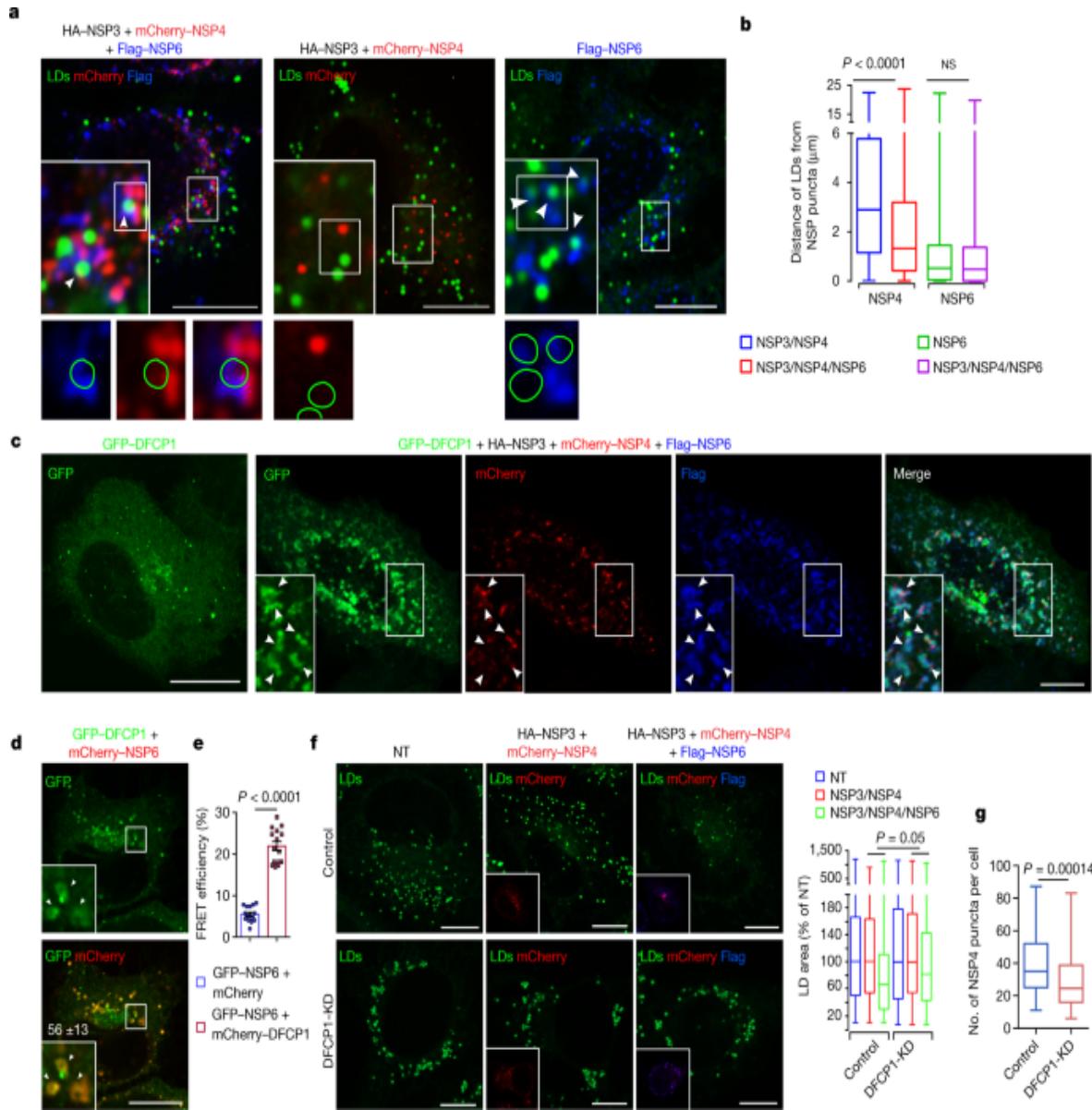
## NSP6 mediates replication organelle–LD association

A C-terminal 80-amino-acid fragment of NSP6 (NSP6-C80), which is unable to induce the formation of the NSP6 compartment, was associated with roundish cytoplasmic structures. These were negative for endosomal, Golgi or mitochondrial markers, but turned out to be lipid droplets (LDs) (Extended Data Fig. 9a,b). This association is due to the amphipathic helix, as a mutated version of NSP6-C80 (NSP6-C80(F220Q/T222W)) that lost the amphiphilic properties of its amphipathic helix did not associate with LDs and exhibited a diffuse distribution (Extended Data Fig. 9c).

Of note, and in agreement with previous results<sup>24</sup>, we found that 40% of the viral replication areas labelled by double-stranded RNA (dsRNA) and NSP6 are associated with LDs (Extended Data Fig. 9d), and that LDs are required for the replication of SARS-CoV-2 in Calu-3 cells because A922500—an inhibitor of DGAT-1—inhibited LD biogenesis and significantly reduced the viral load (Extended Data Fig. 9e).

We found that LDs were also in close proximity to ROLS in cells that co-expressed NSP3/NSP4/NSP6, mimicking the situation of infected cells, but not in cells that expressed only NSP3/NSP4 (Fig. 4a,b and Extended Data Fig. 9f,g). By contrast, LDs were found very close to NSP6 structures in cells that expressed NSP6 alone (Fig. 4a,b and Extended Data Fig. 9g). These data indicate that NSP6 mediates the association of LDs with ROLS.

**Fig. 4: NSP6 mediates the recruitment of LDs to the replication organelle through DFCP1.**



**a**, HeLa cells expressing the indicated NSPs stained with BODIPY-488 for LDs (green). Insets, enlargement of boxed area; arrowheads, LDs close to ROLs (left) or to NSP6 (right). Bottom panels, green circles delineate the position of LDs. **b**, Distance of LDs from NSP4 and NSP6 puncta measured in whole cells expressing the indicated NSPs (see [Methods](#)). LD to NSP4 puncta,  $N=3$ ;  $n=1,692$  in NSP3/NSP4-expressing and  $n=2,971$  in NSP3/NSP4/NSP6-expressing cells. LD to NSP6 puncta,  $N=3$ ;  $n=3,239$  in NSP6-expressing and  $n=2,563$  in NSP3/NSP4/NSP6-expressing cells. **c,d**, HeLa cells expressing GFP-DFCP1 alone or with mCherry-NSP4 + HA-NSP3 + Flag-NSP6 (**c**), or with mCherry-NSP6 (**d**). Arrowheads, DFCP1

signal in the NSP6 compartment. In **d**, the percentage of co-localization between DFCP1 and NSP6 is indicated (mean  $\pm$  s.d.,  $N = 3$ ,  $n = 30$ ). **e**, FLIM–FRET analysis showing the average GFP lifetime in HeLa cells expressing GFP–NSP6 with mCherry or with mCherry–DFCP1. Mean  $\pm$  s.d.,  $N = 3$ ,  $n = 15$ . **f**, LD staining (BODIPY-488, green) of control (mock-transfected) and DFCP1-knockdown (DFCP1-KD) cells expressing mCherry–NSP4 + HA–NSP3, mCherry–NSP4 + HA–NSP3 + Flag–NSP6 or non-transfected (NT). Insets, mCherry–NSP4 fluorescence (red) and anti-Flag immunostaining (blue). Graph, quantification of LD area in cells. Values are normalized to the NT-cells in either the control or the DFCP1-KD condition. The significance of LD area reduction after DFCP1-KD and mock treatment was assessed as described in the [Methods](#).  $N = 3$ ,  $n = 90$ . **g**, Number of NSP4 puncta per cell in mCherry–NSP4 + HA–NSP3 + Flag–NSP6-transfected cells without (control) or with DFCP1-KD.  $N = 3$ ; control  $n = 66$ , DFCP1-KD  $n = 71$ . Scale bars, 10  $\mu$ m (**a,c,d,f**). Kruskal–Wallis test with Wilcoxon post-hoc and Bonferroni correction (**b**); unpaired two-tailed *t*-test (**e**); Wilcoxon test (**f**); unpaired two-tailed Mann–Whitney test (**g**). NS, not significant. Box plots in **b,f,g** represent 25th to 75th percentiles of the data (centre line, median; whiskers, minima and maxima).

## [Source data](#)

We investigated the involvement of molecular complexes that are known to tether LDs to the ER<sup>25,26,27</sup> and found that DFCP1 (Fig. [4c](#)) and RAB18 (Extended Data Fig. [9h](#)) were associated with ROLs. In particular, DFCP1 was recruited by NSP6 but not by NSP3/NSP4 (Fig. [4d](#) and Extended Data Fig. [9i](#)). We found that the two proteins interact, as shown by the intense FRET signal that was measured in cells expressing GFP–NSP6 and mCherry–DFCP1 (Fig. [4e](#)) and by the ability of DFCP1 to pull down NSP6 from the lysates of cells expressing HA–NSP6 (Extended Data Fig. [9j](#)). The C-terminal domain of NSP6 mediates the recruitment of DFCP1, as NSP6(1–157) was unable to recruit DFCP1 (Extended Data Fig. [10a](#)). A DFCP1 mutant (DFCP1( $\Delta$ 1–416)) that lacks the N-terminal domain but includes the ER-targeting domain and the two FYVE domains<sup>28</sup> is still recruited by NSP6 (Extended Data Fig. [10b](#)), as is the FYVE domain mutant C654S/C770S, which is unable to bind phosphatidylinositol 3-phosphate (PtdIns3P) (but not the single-point mutant W543A in the ER domain) (Extended Data Fig. [10b](#)). Thus, unlike recruitment to the omegasome (the

site of autophagosome formation)<sup>29</sup>, recruitment of DFCP1 to the ROLS is independent of PtdIns3P. Indeed, inhibition of PtdIns3P generation, by wortmannin or SAR405, did not affect the recruitment of DFCP1 to the NSP6 compartment (Extended Data Fig. [10c](#)). SAR405 did not impair the formation of the NSP6 compartment, which argues against a role for PtdIns3P in this process. Supporting an autophagy-independent role of DFCP1 recruitment by NSP6, the number of autophagosomes in cells expressing NSP6 was comparable to that of non-transfected cells (Extended Data Fig. [10d](#)).

Notably, we found that LDs are consumed during the formation of ROLS in cells expressing NSP3/NSP4/NSP6, but not in cells expressing NSP3/NSP4 (Fig. [4f](#)), and that a fluorescent fatty acid incorporated into LDs<sup>30</sup> shows more efficient transfer to NSP3/NSP4 structures in the presence of NSP6 (Extended Data Fig. [10e](#))—consistent with a role for NSP6 in channelling LD-derived lipids to the ROLS. Of note, the NSP6-dependent consumption of LDs and the formation of ROLS were both inhibited by the knockdown of DFCP1 (Fig. [4f,g](#) and Extended Data Fig. [10f](#)). Finally, and in line with recent reports<sup>19</sup>, depletion of DFCP1 also inhibited the replication of SARS-CoV-2, confirming that the availability of LDs is required to sustain viral replication (Extended Data Fig. [10g,h](#)).

## Conclusions

The SARS-CoV-2 replication organelle is made of DMVs and connectors<sup>2,8</sup>, the molecular determinants of which we have shown are constituted by NSP3/NSP4 and NSP6, respectively. The NSP6 connectors are tracts of zippered ER that are not accessible to luminal ER proteins or ER membrane proteins with bulky luminal domains, but which are freely accessible to lipids (Extended Data Fig. [11](#)). In addition to linking the DMVs to the ER, the connectors mediate the association of replication organelles with LDs (Extended Data Fig. [11](#)); this is likely to be a dynamic association (as at any given time, 40% of replication organelles are associated with LDs) that could provide fatty acids to fuel DMV growth. These features are perfectly suited to refurbish the DMVs with lipids synthesized in the ER but to exclude 'undesired' ER proteins.

We found that NSP6 zippers the ER membrane through homodimerization and that NSP6(ΔSGF) (which underwent convergent evolution in the Alpha, Beta, Gamma, Eta, Iota and Lambda VOCs) has a higher ER-zippering activity. Notably, the recent and highly infectious BA.2 variant also contains the NSP6(ΔSGF) deletion<sup>18</sup>. The deletion is found in the second and longest luminal loop of NSP6, hosting a consensus *O*-glycosylation motif (LSGF: 105–108), which could act as a spacer that forms luminal bridges. Thus, the SGF deletion could convey higher zippering activity by shortening the ‘spacer’ and/or by preventing its *O*-glycosylation.

Our findings on NSP6 and its key role in replication organelle biogenesis provide a testable target that is easily amenable to screens for antiviral agents and that is applicable to a wide range of coronaviruses.

## Methods

### Reagents and antibodies

The following antibodies were used: mouse monoclonal anti-HA (BioLegend, 901503, dilution 1:600 for immunofluorescence (IF) and 1:1,500 for western blot (WB)), rabbit polyclonal anti-HA (Sigma-Aldrich, H6908, dilution 1:200 for IF), goat polyclonal anti-HA (Bethyl, A190-138A, dilution 1:600 for IF), rabbit polyclonal anti-actin (Sigma-Aldrich, A2066, dilution 1:10,000 for WB), rabbit polyclonal anti-NSP6 (ProSci, 9177, dilution 1:200 for IF and 1:1,000 for WB), sheep anti-NSP3 (The University of Dundee, DA126, dilution 1:100 for IF and 1:1,000 for WB), rabbit polyclonal ADRP/Perilipin 2 (Proteintech, 15294-1-AP, dilution 1:200), rabbit monoclonal anti-DFCP1 (Cell Signaling, 38419, dilution 1:1,000 for WB), mouse monoclonal anti-Flag (Sigma-Aldrich, F1804, dilution 1:400 for IF and 1:1,500 for WB), goat polyclonal anti-Flag (Bethyl, A190-101A, dilution 1:200 for IF), mouse monoclonal anti-c-Myc (Santa Cruz, sc-40, dilution 1:200 for IF), mouse monoclonal anti-GAPDH (Santa Cruz, sc-32233, dilution 1:1,000 for WB), mouse monoclonal anti-LAMP1 (Hybridoma Bank, H4A3, dilution 1:200 for IF), rabbit monoclonal anti-EEA1 (BD Biosciences, 610456, dilution 1:1,000 for IF), sheep anti-human anti-TGN46 (BioRad, AHP500GT, dilution 1:750 for IF), rabbit polyclonal anti-GFP (Abcam, ab6556, dilution 1:250 for IF), mouse monoclonal anti-

GFP (Santa Cruz, sc-9996, dilution 1:2,000 for WB), mouse monoclonal anti-mCherry (Abcam, ab125096, dilution 1:2,000 for WB), mouse monoclonal anti-V5 (Thermo Fisher Scientific, R960-25, dilution 1:200 for IF and 1:1,000 for WB), rabbit polyclonal anti-LC3 (Novus Biologicals, NB100-2220, dilution 1:200 for IF), mouse monoclonal anti-dsRNA (Scicons, 10010500, dilution 1:10 for IF), DAPI (Sigma-Aldrich, D9542, dilution 1:10,000 for IF), rabbit 1.4-nm gold-conjugated Fab' fragment (Nanoprobes, 2004, dilution 1:50), mouse 1.4-nm gold-conjugated Fab' fragment (Nanoprobes, 2002, dilution 1:50), Alexa Fluor-546 FluoroNanogold anti-mouse Fab' (7402, dilution 1:50) and Alexa Fluor-488-568-647 (Invitrogen, diluted 1:400), horseradish peroxidase (HRP)-conjugated goat anti-mouse or anti-rabbit IgG antibody (1:8,000, Merck Millipore, 401215 or 401315, respectively). Anti-GM130 (1:1,000 for IF) and anti-VAP-A (1:300 for IF) were produced in our laboratory as previously described<sup>[31](#),<sup>[32](#)</sup></sup>

BODIPY 493/503 (4,4-difluoro-1,3,5,7,8-pentamethyl-4-bora-3a,4a-diaza-s-indacene), β-BODIPY FL C<sub>12</sub>-HPC (2-(4,4-difluoro-5,7-dimethyl-4-bora-3a,4a-diaza-s-indacene-3-dodecanoyl)-1-hexadecanoyl-sn-glycero-3-phosphocholine) and BODIPY 558/568-DA-C<sub>12</sub> (4,4-difluoro-5-(2-thienyl)-4-bora-3a,4a-diaza-s-indacene-3-dodecanoic acid were purchased from Thermo Fisher Scientific (D3922, D3792 and D3835, respectively). Oil Red O solution was purchased from Merck (102419). K22 (N-[(1Z)-1-[[4-(4-bromophenyl)-4-hydroxy-1-piperidinyl]carbonyl]-2-phenylethenyl]-benzamide) was purchased from Cayman Chemical, the DGAT-1 inhibitor A922500 (A1737), wortmannin (3144), delipidated serum (S5394) and doxycycline hydrochloride (8D3447) from Sigma-Aldrich and the VPS34 specific inhibitor SAR405 from MedChemExpress (HY-12481). Puromycin dihydrochloride was purchased from Calbiochem (540411). For <sup>35</sup>S-methionine/cysteine labelling, the EasyTag protein labelling mix (772007MC) was purchased from PerkinElmer. Unless otherwise stated, all other chemicals were purchased from Sigma-Aldrich.

## Plasmid constructs

All NSP constructs were made with the Gateway system (Thermo Fisher Scientific) using a modified pCDNA3.1 vector (containing a HA, Flag, Myc,

GFP or mCherry tag) for amino-terminal tagging, a modified pCDNA5/FRT/TO vector (containing 3×Flag) for carboxy-terminal tagging, unmodified pCDNA5/FRT/TO to clone untagged NSP6, and pLTD-Flag or pLTD-HA for stable doxycycline-inducible NSP6-expressing cell lines. All Gateway vectors were provided by P. Grumati. The donor plasmids were pDONR207 SARS-CoV-2 NSP3, pDONR223 SARS-CoV-2 NSP4 and pDONR223 SARS-CoV-2 NSP6 from Wuhan-Hu-1 SARS-CoV-2 (gifts from F. Roth, Addgene plasmids 141257, 141258 and 141260, respectively)<sup>33</sup>. For carboxy-terminal tagging of NSP6, the stop codon was removed using the oligo pairs NSP6 ns(+)/NSP6 ns(−) (Supplementary Table 2) with the Agilent QuikChange kit. The Agilent QuikChange kit and the oligos described in Supplementary Table 2 were used to make the following NSP6 N-terminally tagged mutant constructs: NSP6(1–157) (amino acids 1–157); NSP6-C80 (amino acids 211–290); the mutants in the amphiphilic alpha helix NSP6(F220Q/T222W) and NSP6-C80(F220Q/T222W); and the VOC mutant constructs NSP6(ΔSGF) and NSP6(ΔSGF)–NSP7.

The NSP6–NSP7 sequence was synthesized with flanking attB sequences by Thermo Fisher Scientific (Supplementary Table 1), a V5 tag was added to NSP7 by PCR, and the amplicon was cloned into the Gateway vector pDONR223 and recombined with destination vector pCDNA3.1 containing HA to produce pH-A-NSP6-NSP7-V5.

The IBV (avian infectious bronchitis virus, strain M41) NSP6 sequence (corresponding to Uniprot P0C6Y3 from position 3089 to 3381), optimized for human expression and synthesized with flanking attB sequences by Thermo Fisher Scientific (Supplementary Table 1), was cloned into the Gateway vector pDONR223 and recombined in Flag-containing Gateway destination vectors at the amino or carboxy terminus. Oligos NSP6-IBV ns(+)/NSP6-IBV ns(−) (Supplementary Table 2) were used to remove the stop codon for the carboxy terminal-tagged construct.

mCherry-DFCP1 was a gift from D.-H. Kim (Addgene plasmid 86746). pEGFP-ATF6 was a gift from R. Prywes (Addgene plasmid 32955). mCherry-Calreticulin-N-16 (M. Davidson, Addgene plasmid 55006), pLenti-X1-Neo-GFP-ATL2 (J. Corn, Addgene plasmid 109020), pEGFPC-

DFCP1 and pRUBY-N1-KDEL were provided by P. Grumati. pEGFP-RAB18 was a gift from M. Scidmore (Addgene plasmid 4955).

The Agilent QuikChange kit and the oligos described in Supplementary Table 2 were used to make the following mCherry–DFCP1 mutant constructs: DFCP1(Δ1–416) (lacking the amino terminus); DFCP1(W543A) (point mutation in the ER domain) and DFCP1(C654S/C770S) (mutations in the double FYVE domain; unable to bind PtdIns3P).

GST-tagged DFCP1 was constructed by amplifying the coding sequence from mCherry–DFCP1 with oligos DFCP1-p223(+) / DFCP1-p223(–) and cloning into the Gateway vector pDONR223 and subsequently into the Gateway vector pET60.

pEYFPC3-Cb5, constructed as described<sup>12</sup> using YFP instead of mCherry, and pEGFP-VAP-A were made in our laboratory. pEGFP-ERGIC53 and p-KDELR-EGFP were gifts from A. Luini.

BP clonase and LR clonase for Gateway cloning were purchased from Thermo Fisher Scientific. All other reagents for molecular biology were purchased from New England Biolabs.

## Cell culture, transfection and RNA interference

HeLa cells were obtained from ATCC and cultured as previously described<sup>12</sup>. Calu-3 cells (human lung adenocarcinoma), a gift from L. J. Galietta, were cultured in DMEM F-12 (Gibco), supplemented with 10% fetal bovine serum (Euroclone) 100 IU ml<sup>-1</sup> penicillin and 100 µg ml<sup>-1</sup> streptomycin (Thermo Fisher Scientific) and 2 mM l-glutamine (Thermo Fisher Scientific) in a humidified incubator at 37 °C and 5% CO<sub>2</sub>. Cell lines were routinely tested for mycoplasma (Biological Industries). Cells were transfected with plasmids using either TransIT-LT1 (Mirus Bio) for HeLa cells or Lipofectamine LTX and PLUS Reagent (Thermo Fisher Scientific) for Calu-3 according to the manufacturer's instructions. Expression was maintained for 16–24 h before processing unless otherwise stated. For RNA interference, HeLa and Calu-3 cells were mock-treated or treated with DFCP1 siRNA (50 nM) for 96 h using Lipofectamine RNAiMAX (Thermo

Fisher Scientific) for direct transfection. The siRNA sequences used in this study are listed in Supplementary Table 2.

## Generation of HeLa Flag–NSP6 and HA–NSP6 doxycycline-inducible stable lines

To generate stably expressing clones, HeLa cells were transfected with the plasmids pLTD-Flag-NSP6, pLTD-Flag-NSP6( $\Delta$ SGF), pLTD-HA-NSP6 or pLTD-HA-NSP6( $\Delta$ SGF) and selected with complete medium containing 3  $\mu\text{g ml}^{-1}$  puromycin (Calbiochem). Single-cell cultures were isolated from the mixed populations and protein expression was probed and induced with 1  $\mu\text{g ml}^{-1}$  doxycycline (Sigma-Aldrich) at different time points, as indicated. Samples were then processed by immunofluorescence analysis. All the cell lines generated in this study were authenticated through western blot and immunofluorescence.

## SARS-CoV-2 infection and assays

SARS-CoV-2 infection, virus titration and cell death assay through the activity of lactate dehydrogenase (LDH) were performed as elsewhere described<sup>24</sup>. For immunofluorescence experiments, Calu-3 cells were seeded on coverslips, left untreated or pre-treated for 2 h with K22 or with the DGAT-1 inhibitor A922500 at different concentrations, as indicated in the figures. Cell number and cell viability after treatment with either K22 or A922500 were assessed by crystal violet staining, cell morphology analysis or LDH assay. No cytostatic or cytotoxic effect of the drugs was observed at the concentrations used. For immunofluorescence experiments and drug treatments, Calu-3 cells were seeded on coverslips and infected with SARS-CoV-2 early lineage (SARS-CoV-2/human/BRA/RJ01/2020, GenBank accession no. MT710714) at a multiplicity of infection (MOI) of 0.01 for 48 h. Infected cells were fixed with 3.7% formaldehyde and processed for immunofluorescence as described<sup>24</sup>. For comparative analyses of NSP3–NSP6 proximity, cells were similarly infected with early lineage and Gamma variant (hCoV-19/Brazil/AM-L70-71-CD1739/2020, GISAID ID: EPI\_ISL\_1060902) SARS-CoV-2 at a MOI of 0.01 for 48 h.

For EM experiments, Calu-3 cells were infected with early lineage B.1 (hCoV-19/Italy/CAM-INMI-32803-66/2020, GISAID ID: EPI\_ISL\_493333) or Gamma variant (hCoV-19/Italy/CAM-IZSM-RD020483D54/2021, GISAID ID: EPI\_ISL\_2933105) SARS-CoV-2 strains at an MOI of 10 for 24 h. SARS-CoV-2-infected Calu-3 cells were processed for EM as described below. All procedures related to virus culture were handled at a biosafety level 3 (BSL3) multi-user facility, according to World Health Organization (WHO) guidelines.

## Drug treatments

Flag–NSP6- and mCherry–DFCP1-transfected cells were treated with either 100 nM wortmannin or 1  $\mu$ M VPS34 inhibitor SAR405 for 3 h, then processed for immunofluorescence. For K22 treatment, cells were transfected and after 30 min dimethyl sulfoxide (DMSO) or 40  $\mu$ M K22 were added.

## Recombinant proteins and pull-down and co-immunoprecipitation experiments

All recombinant proteins were purified from *Escherichia coli* Rosetta DE3 cells (Merck). GST-tagged DFCP1 from plasmid pET60 and GST alone from plasmid GEX-4T2 (GE Healthcare) were expressed as described<sup>34</sup>. For pull-down experiments, 3 mg of cellular lysates from HA-NSP6-transfected HeLa cells was incubated with GST–DFCP1 or GST alone (0.1  $\mu$ M) overnight at 4 °C in 950  $\mu$ l binding buffer (25 mM Tris pH 7.4, 150 mM NaCl, 0.1% Triton-X-100, 0.1% NP-40, 1 mM EDTA and protease inhibitors). Glutathione beads were added, incubated for 1 h at 4 °C, washed four times with incubation buffer and twice with a similar buffer without detergents, eluted and analysed by SDS–PAGE.

For co-immunoprecipitation experiments, 1.7 mg of cellular lysate from cells mock-transfected or co-transfected with HA–NSP6 together with GFP–NSP6, Flag–NSP6, GFP–ERGIC53, GFP–atlastin-2 or GFP–NSP6(1–157), or co-transfected with HA–NSP6( $\Delta$ SGF) and GFP–NSP6( $\Delta$ SGF), were incubated with appropriate antibody-conjugated beads (HA, Flag and GFP). After overnight incubation at 4 °C in 750  $\mu$ l binding buffer, samples were

washed five times with binding buffer and once with a similar buffer without detergents, eluted and analysed by SDS-PAGE. To evaluate co-immunoprecipitation efficiency, a total of three independent experiments were analysed. The co-immunoprecipitated GFP-NSP6 signal was divided by the GFP-NSP6 signal in the input and normalized by the signal of the immunoprecipitated primary antigen (HA). Co-immunoprecipitation efficiency was reported as mean  $\pm$  s.e.m. of co-immunoprecipitated GFP-NSP6( $\Delta$ SGF) compared to GFP-NSP6.

## Detergent extraction

HeLa cells transfected with Flag-NSP6, NSP6-Flag or Flag-NSP6( $\Delta$ SGF) were lysed in buffer (25 mM Tris pH 7.4, 150 mM NaCl, 1 mM EDTA with protease and phosphatase inhibitor cocktails) containing increasing concentrations of Triton-X-100 and NP-40 (1:1) and centrifuged at 13,200 rpm for 10 min. The pellet was resuspended in the same volume as the supernatant and equal volumes were subjected to western blot analysis using an anti-Flag antibody.

## Metabolic radiolabelling

For metabolic labelling, wild-type HeLa cells or the pLTD-HA-NSP6 or pLTD-HA-NSP6- $\Delta$ SGF stable cell lines were induced with doxycycline ( $1 \mu\text{g ml}^{-1}$ ) for 13 h, incubated for 30 min with methionine/cysteine-free medium (21013024, Gibco), and then incubated for 1 h at  $37^\circ\text{C}$  with  $50 \mu\text{Ci ml}^{-1}$   $^{35}\text{S}$ -methionine/cysteine (PerkinElmer) in the same medium. The cells were then washed three times with complete medium and further incubated for different times at  $37^\circ\text{C}$  in complete medium. Doxycycline ( $1 \mu\text{g ml}^{-1}$ ) was included in all media. After cell lysis, proteins were immunoprecipitated with anti-HA affinity beads, and analysed by SDS-PAGE gel autoradiography (using a Typhoon Imager, Image QuantTool, GE Healthcare) of the immunoprecipitates to measure protein stability followed by immunoblot using anti-HA to measure total protein levels.

## Western blot analysis

Western blot analysis and densitometry were performed as previously described<sup>34</sup>. Samples containing NSP6 were mixed with sample buffer (100 mM Tris pH 6.8, 25% glycerol, 2% SDS, 0.01% bromophenol blue and 10% 2-mercaptoethanol), but were not boiled before loading.

## Immunofluorescence analysis

Immunofluorescence analysis was performed as previously described<sup>12</sup>.

## Digitonin and Triton-X-100 permeabilization

HeLa cells transfected with Flag-NSP6 or NSP6-Flag were grown on coverslips and fixed with 4% PFA for 10 min, washed three times with buffer A (20 mM PIPES pH 6.8, 137 mM NaCl and 2.7 mM KCl) and permeabilized with 20 µM digitonin (Calbiochem) diluted in buffer A for 5 min. Coverslips were blocked for 30 min with blocking solution (5% FBS (v/v) and 50 mM NH<sub>4</sub>Cl in buffer A) without any additional permeabilizing agent and incubated with primary anti-Flag and anti-TGN46 antibodies diluted in blocking solution. The TGN46 antibody was raised against a luminal portion of the protein that is thus not accessible after digitonin permeabilization. This represents a control in that only the plasma membrane has been permeabilized. Coverslips were washed with buffer A and incubated with fluorochrome-conjugated secondary antibodies (Alexa Fluor-488 for Flag and Alexa Fluor 568 for TGN46 in buffer A) for 1 h at room temperature. After incubation, cells were fixed with 2% PFA for 5 min and washed once with 50 mM NH<sub>4</sub>Cl in PBS. Coverslips were subsequently permeabilized with 0.1% Triton-X-100 in PBS for 5 min. Cells were then blocked with blocking solution (0.05% saponin, 0.5% BSA and 50 mM NH<sub>4</sub>Cl in PBS) and incubated with the same primary antibodies used in the first step. Coverslips were then washed with PBS and incubated with fluorochrome-conjugated secondary antibodies (Alexa Fluor 405 for Flag and Alexa Fluor 633 for TGN46 in PBS) for 1 h at room temperature. The TGN46 epitope becomes accessible to the primary antibody under these conditions, confirming selective permeability and identifying luminal epitopes.

## **LD staining and assays**

LDs were stained by adding 0.5 µM BODIPY 493/503 (Thermo Fisher Scientific) to the fluorochrome-conjugated secondary antibody mix for 30 min after fixation and processed as for immunofluorescence analysis.

To monitor lipid transfer from LDs to DMVs we followed the protocol described previously<sup>30</sup>. In brief, BODIPY 558/568-DA-C<sub>12</sub> at a final concentration of 1 µM was added for 16 h to the culture medium of HeLa cells transfected with GFP–NSP4/HA–NSP3 or GFP–NSP4/HA–NSP3/Flag–NSP6. Cells were then washed and incubated with DMEM supplemented with delipidated serum (1%) for an additional 6 h. Coverslips were fixed and processed as described above. NSP4 puncta were identified by using the ‘Analyze particles’ tool of Fiji (ImageJ) software, and the fluorescence mean intensity of Bodipy-DA-C<sub>12</sub> for each particle was determined. Particles with values equal or higher than a similar area of the ER were defined as ‘positive’ particles. The percentage of NSP4 Bodipy-DA-C<sub>12</sub>-positive particles was calculated for each cell.

## **Confocal microscopy and image analyses**

Cells were imaged using a Plan-APOCHROMAT 100×/1.4 oil objective on a Zeiss LSM800 or LSM880 confocal system equipped with an AiryScan module and controlled by the Zen blue software. Fluorescence images presented are representative of images collected from at least three independent experiments, unless otherwise stated (see 'Statistics and Reproducibility' for further details). The images used for phenotype quantification were acquired with the same parameters (that is, digital gain, laser power and magnification) and processed with Fiji (ImageJ; National Institutes of Health (NIH)) software. Brightness and contrast were adjusted with Adobe Photoshop, and figure panels were assembled with Adobe Illustrator.

## **Quantification of the number and area of structures**

NSP6, NSP4, LC3 and LD structures were analysed using the 'Analyze particle' function to determine their number per cell. For each experiment, images were acquired below saturation limit and the same threshold was chosen and applied to all of them. For the calculation of the size of the structures the 'Analyze particle' function was used, setting 'Area' as measurement.

### **Distribution of NSP4 puncta**

To calculate the distribution of NSP4 puncta in each cell, the 'Analyze particle' function was used, considering a particle size between 0.1 and infinity and choosing the centre of mass as a reference for measurement. *X* and *Y* coordinates for each NSP4 puncta were obtained and plotted. A four-quadrant subdivision was applied to the images using the *XY* coordinates of the centre of mass as the axis origin. The relative abundance of the NSP4 puncta for each quadrant is expressed as a percentage of the total identified structures for each cell.

### **Relative distribution of the NSP6 protein**

To measure the cellular distribution of NSP6 fluorescence, the integrated density of NSP6 in NSP6 structures was calculated over the integrated density of total NSP6 in the whole cell.

Cells with comparable levels of total integrated fluorescence intensity were analysed for each time point. Results were expressed as a percentage of the fluorescent NSP6 signal present in the NSP6 structures over the total fluorescence.

### **Recruitment to NSP6 structures**

The fraction of VAP-A or NSP6(1–157) associated with NSP6-positive structures was measured as the ratio between the integrated density of each protein on the NSP6 structures and the integrated density in the whole cell.

### **Co-localization between NSP6 and DFCP1**

Co-localization between NSP6 and wild-type or mutant DFCP1 was calculated using the JACoP plug-in<sup>35</sup>.

## Distance between particles

The relative distance between objects was determined with the DiAna plug-in<sup>36</sup>. In brief, channels were thresholded and then segmented. For LD distance from NSP4 and NSP6 in transfected cells shown in Fig. 4b, edge–edge distances between particles were measured in the whole cell. No values were excluded. In addition, for selected images including the one in Fig. 4a we applied the Shuffle function<sup>36</sup> as shown in Extended Data Fig. 9f. In brief, this function redistributes the objects in a channel in a random manner; then the distances between objects of the randomized channel to the closest object in the second channel from the original image are measured. The distribution of these distances is represented as the mean (red line) flanked by 95% confidence intervals (green lines). The distribution of the distances measured between the objects in the two channels from the original images is plotted (blue line). If this distribution falls outside the confidence interval of the distance obtained for shuffled images, the distance is considered as statistically significant ( $P < 0.05$ ).

For the proximity between LDs and dsRNA or NSP6 in Extended Data Fig. 9d, and the proximity between dsRNA and NSP6 in Extended Data Fig. 7i, in infected cells, the edge–edge distance was analysed and structures closer than 250 nm (in all directions) were considered as associated structures. To calculate the distance between NSP3- and NSP6-positive structures in infected cells in Fig. 3b both centre–centre and edge–edge distances were measured.

## Measurements of NSP6 fluorescence intensity

HeLa cells expressing Flag-tagged NSP6 were fixed and processed for immunofluorescence. Cells with similar expression were acquired using the same parameters and processed with the Fiji (ImageJ) software. The integrated density of each cell was measured.

## Electron microscopy

For pre-embedding IEM, the cells were fixed, permeabilized and labelled as described previously<sup>37</sup>. In brief, the cells were fixed with a mixture of 4% paraformaldehyde (PFA) and 0.05% glutaraldehyde prepared in 0.2 M HEPES buffer for 10 min (room temperature) and then with 4% PFA alone for 30 min (room temperature), followed by incubation with blocking/permeabilizing solution (0.5% bovine serum albumin (BSA), 0.1% saponin and 50 mM NH<sub>4</sub>Cl in PBS) for 30 min.

Cells were incubated with a primary anti-HA monoclonal antibody (1:600, BioLegend) diluted in blocking/permeabilizing solution overnight and then a secondary anti-mouse antibody (1.4-nm gold-conjugated Fab' fragment diluted 1:50, Nanoprobes) was added for 2 h. The GoldEnhance EM kit (from Nanoprobes) was used to enhance ultrasmall gold particles. For double labelling of cells expressing HA–NSP3, mCherry–NSP4 and GFP–NSP6, enhancement with the anti-HA antibody was performed for 3 min and then a primary anti-GFP polyclonal rabbit antibody (1:250, Abcam) was added and processed as above using a secondary anti-rabbit antibody (1.4-nm gold-conjugated Fab' fragment diluted 1:50, Nanoprobes) for 2 h, followed by gold enhancement for an additional 3 min. The longer enhancement time for the anti-HA detection causes the formation of larger gold particles (clusters) with an irregular shape that distinguishes HA–NSP3 from the smaller GFP–NSP6 signals in doubly transfected cells.

For conventional EM, the cells were fixed with 1% GA prepared in 0.2 M HEPES buffer for 30 min (RT).

Cells prepared for IEM or conventional EM were scraped, pelleted, post-fixed in OsO<sub>4</sub> and uranyl acetate, dehydrated, embedded in Epon and polymerized at 60 °C for 72 h. For each sample, thin sections were cut using a Leica EM UC7 ultramicrotome (Leica Microsystems). EM images were acquired from thin sections using a FEI Tecnai-12 electron microscope (FEI) equipped with a VELETTA CCD digital camera (Soft Imaging Systems). Morphometric analysis of the structures of interest was performed using iTEM software (Olympus).

## CLEM

HeLa cells were transfected with HA–NSP6 or HA–NSP6( $\Delta$ SGF), or they were co-transfected with HA–NSP3/mCherry–NSP4/GFP–NSP6 or HA–NSP3/mCherry–NSP4/Myc–NSP6 where indicated. Transfected cells were treated or not with 40  $\mu$ M K22 30 min after transfection. After overnight expression, cells were fixed as for IEM and then labelled with an anti-HA antibody followed by detection with a secondary Alexa Fluor-546 FluoroNanogold anti-mouse Fab'. The structures of interest carrying different proteins were visualized by confocal microscopy using a Zeiss LSM800 station and fluorescent images were recorded. Then the cells were post-fixed, dehydrated, embedded in Epon and polymerized as described above. Serial 60-nm sections were cut and analysed using a FEI Tecnai-12 electron microscope. The same cell and structures of interest obtained by confocal microscopy were identified on EM images using Zen Connect software (Zeiss).

## Electron tomography

Epon sections (250 nm thick) were collected on Formvar carbon-coated slot grids and analysed using a Tecnai G2 Spirit BioTwin electron microscope (FEI) equipped with an automated tomography stage. The single tilt series of images were acquired in a range of  $-65^\circ$  to  $+65^\circ$  (at  $1^\circ$  intervals) using Xplore 3D TEM Tomography software (FEI) at 40,000 $\times$  magnification unless otherwise stated. Tilt series were used with the open-source IMOD software to generate tomograms. At least 10 tomograms were analysed per experimental condition. For 3D reconstruction, the surfaces of DMVs and surrounding ER membranes were rendered using the IMOD software.

## FLIM measurements, FRAP and FLIP analysis

FLIM–FRET analysis of GFP–NSP6 alone and in combination with mCherry–NSP6, mCherry or mCherry–DFCP1, and of mCherry–NSP6 with GFP–NSP6(1–157), GFP–ERGIC and GFP–atlastin-2, was performed as previously described<sup>12</sup>. FLIM data analysis was performed using SymPhoTime 64 (Picoquant). For live-cell imaging of the NSP6 structures, cells were plated in glass-bottomed dishes (MatTek), transfected with the fluorescently tagged protein constructs or incubated with  $\beta$ -BODIPY FL C<sub>12</sub>-HPC (1  $\mu$ M) for 16 h, and imaged with an LSM800 microscope (Zeiss)

fitted with 488- and 561-nm argon laser lines, using a  $63\times$  Plan-APOCHROMAT NA 1.4 DIC oil immersion objective. During imaging, cells were maintained in complete culture medium in a humidified atmosphere at 37 °C.

Fluorescence images presented are representative of cells imaged in at least three independent experiments and were processed with Fiji (ImageJ; NIH) software.

FRAP experiments and time-lapse laser-scanning confocal microscopy were performed as described<sup>12</sup>. In brief, a single NSP6 structure was acquired five frames before bleaching (6 s per frame). Bleaching was performed with 100% power of the 488 laser for 10 iterations. Recovery was monitored for 600 s after the bleaching event. At least 30 independent structures were analysed for each condition in three different experiments. Data were exported using Zen software (Zeiss) and corrected for bleaching by dividing the fluorescence intensity of the bleached area by that of an unbleached area. Bleaching was minimal during the time course of recovery (between 0 and 10%): where bleaching exceeded 10%, the recovery sequences were discarded. Quantification of GFP–NSP6 and GFP–NSP6( $\Delta$ SGF) dissociation from membranes was measured in living cells by FLIP. FLIP was performed in cells expressing each GFP-tagged protein by bleaching iteratively (100 times, with intervals of 6 s between frames) the GFP-associated fluorescence in the entire cell area except for a region of interest (ROI) containing NSP6 structures. The ROI usually accounted for 10–15% of the total cell area. The relative fluorescence intensity of single structures expressed as a percentage of pre-bleaching fluorescence was plotted as mean values  $\pm$  s.d. A slowdown of the FLIP-induced decay curves of GFP–NSP6( $\Delta$ SGF) from the structures was observed indicating an increase in GFP–NSP6( $\Delta$ SGF) association with membranes.

## EM quantification

The percentage of normal and zippered ER (or NE) surface was quantified in random thin sections from pellets of NSP6-transfected HeLa cells using morphometric grids with the iTEM software (Olympus SIS). Quantification of gold particles in thin sections from HeLa cells expressing HA–NSP6 or HA–NSP6( $\Delta$ SGF) and immuno-gold labelled for HA was performed with the touch count tool of the iTEM software. This quantification was further

used as a measure of HA–NSP6 or HA–NSP6( $\Delta$ SGF) expression in each analysed cell to normalize the surface area of zippered ER for the expression level of the corresponding HA-tagged NSP6 protein. To assess the effect of NSP6 or NSP6( $\Delta$ SGF) on the organization of DMVs, tomograms of DMV clusters were used to quantify the following parameters: DMV diameter, shape factor (ratio between long and short axes), density (number per DMV cluster area), length of ER–DMV connections, number of DMVs per connection and overall number of ER–DMV connections per DMV cluster. DMV cluster was defined as a group of DMVs whose distance from the nearest neighbour does not exceed two average DMV diameters. All measurements in tomograms were done with the 3D Manager plug-in of the open-source Fiji software. The same tools were used to quantify the length of zippered DMV connectors in tomograms from Calu-3 cells infected with the early lineage B.1 or Gamma variant of SARS-CoV-2.

## NSP6 protein topology

NSP6 topology modelling was performed using the Constrained Consensus TOPology prediction server (CCTOP, Institute of Enzymology). The amphipathic features of the  $\alpha$ -helix were determined using HELIQUEST (<http://heliquest.ipmc.cnrs.fr>)<sup>16</sup> and the mutations were introduced following the Genetic Algorithm-based module. The images and cartoons shown in Fig. 2a and Extended Data Fig. 11 were created with BioRender.com.

## Phylogenetic analysis

The phylogenetic analysis of SARS-CoV-2 genomes deposited in the GISAID database (<https://www.gisaid.org/>) was performed on a set of 3,508 representative genomes sampled from December 2019 to July 2021, provided by Nextstrain<sup>18</sup> (<https://nextstrain.org/ncov/global>). The percentages of genomes carrying the SGF deletion in the NSP6 protein were evaluated on samples deposited at GISAID up to 16 July 2021.

## Statistics and reproducibility

Statistical analyses were performed using GraphPad Prism7 (GraphPad Software) or the R software environment for statistical computing (rstatix R

package).

To test the normal distribution of the data and the homogeneity of variance across groups, the Shapiro–Wilk test and Levene’s test were used on the ANOVA residuals. When measured variables were normally distributed, the statistical significance of difference in measured variables between control and treated groups was determined by *t*-test or ANOVA followed by appropriate multiple comparison post-hoc tests depending on the experiment. When the measured variables were not normally distributed, non-parametric Mann–Whitney or Kruskal–Wallis tests were performed followed by appropriate multiple comparison post-hoc tests depending on the experiment.

All the experiments for which statistics was derived were performed three times with similar results; *N* indicates the number of experiments and *n* the number of total measurements or observations. All of the replicates performed were biological and not technical. Detailed information for each experiment is provided below.

The experiments shown in Extended Data Figs. [1a,e](#), [3h](#), [6a](#) and [9c](#) were repeated twice.

The experiments shown in Figs. [1b,c,g–i](#), [2i,j](#), [3d,e,f,i,n](#) and [4c](#) and Extended Data Figs. [1f–j](#), [2a–c](#), [3a,d](#), [4c–i](#), [5k,l](#), [6d,e,i,j,l,m](#), [7c–e](#), [8a,b,f–h](#), [9a,b,f–h](#) and [10a,b](#) were repeated three times.

The experiments shown in Extended Data Figs. [1c,d](#), [2d,e](#), [5c,m](#) and [6c,f](#) were repeated four times. The experiments shown in Figs. [1a](#), [3a](#) and [4d](#) and Extended Data Figs. [6g](#), [9i](#) and [10c](#) were repeated five times. The experiment shown in Extended Data Fig. [5b](#) was repeated six times.

The experiments shown in Figs. [1e,f](#) and [3c](#) and Extended Data Figs. [5j](#) and [6b](#) were repeated ten times.

## Reporting summary

Further information on research design is available in the [Nature Research Reporting Summary](#) linked to this paper.

## Data availability

Full scans for all western blots and autoradiographs are provided in Supplementary Fig. 1. The nucleotide sequences of synthetic IBV NSP6 and NSP6–NSP7 used in this study are provided in Supplementary Table 1. The oligonucleotides, siRNAs and primers used in this study are provided in Supplementary Table 2. Raw data supporting the findings of this study are deposited in Zenodo and are publicly available at <https://doi.org/10.5281/zenodo.5929088>. Raw EM data, including tilt series and reconstructed 3D tomograms, were deposited in the Electron Microscopy Data Bank (EMDB) and the Electron Microscopy Public Image Archive (EMPIAR) public databases with accession codes [EMD-14179](#) and [EMPIAR-10935](#), respectively. [Source data](#) are provided with this paper.

## References

1. Hartenian, E. et al. The molecular virology of coronaviruses. *J. Biol. Chem.* **295**, 12910–12934 (2020).
2. Cortese, M. et al. Integrative imaging reveals SARS-CoV-2-induced reshaping of subcellular morphologies. *Cell Host Microbe* **28**, 853–866.e5 (2020).
3. Angelini, M. M., Akhlaghpour, M., Neuman, B. W. & Buchmeier, M. J. Severe acute respiratory syndrome coronavirus nonstructural proteins 3, 4, and 6 induce double-membrane vesicles. *mBio* **4**, e00524–13 (2013).
4. Oudshoorn, D. et al. Expression and cleavage of Middle East respiratory syndrome coronavirus nsp3-4 polyprotein induce the formation of double-membrane vesicles that mimic those associated with coronaviral RNA replication. *mBio* **8**, e01658-17 (2017).
5. Wolff, G., Melia, C. E., Snijder, E. J. & Bárcena, M. Double-membrane vesicles as platforms for viral replication. *Trends Microbiol.* **28**, 1022–1033 (2020).

6. Snijder, E. J. et al. A unifying structural and functional model of the coronavirus replication organelle: tracking down RNA synthesis. *PLoS Biol.* **18**, e3000715 (2020).
7. Klein, S. et al. SARS-CoV-2 structure and replication characterized by *in situ* cryo-electron tomography. *Nat. Commun.* **11**, 5885 (2020).
8. Ogando, N. S. et al. SARS-coronavirus-2 replication in Vero E6 cells: replication kinetics, rapid adaptation and cytopathology. *J. Gen. Virol.* **101**, 925–940 (2020).
9. Peacock, T. P., Penrice-Randal, R., Hiscox, J. A. & Barclay, W. S. SARS-CoV-2 one year on: evidence for ongoing viral adaptation. *J. Gen. Virol.* **102**, 001584 (2021).
10. Romero, P. D. et al. The emergence of SARS-CoV-2 variant lambda (C.37) in South America. *Microbiol. Spectr.* **9**, e0078921 (2021).
11. Cottam, E. M. et al. Coronavirus nsp6 proteins generate autophagosomes from the endoplasmic reticulum via an omegasome intermediate. *Autophagy* **7**, 1335–1347 (2011).
12. Venditti, R. et al. Molecular determinants of ER–Golgi contacts identified through a new FRET-FLIM system. *J. Cell Biol.* **218**, 1055–1065 (2019).
13. Dobson, L., Reményi, I. & Tusnády, G. E. CCTOP: a Consensus Constrained TOPology prediction web server. *Nucleic Acids Res.* **43**, W408–W412 (2015).
14. Baliji, S., Cammer, S. A., Sobral, B. & Baker, S. C. Detection of nonstructural protein 6 in murine coronavirus-infected cells and analysis of the transmembrane topology by using bioinformatics and molecular approaches. *J. Virol.* **83**, 6957–6962 (2009).
15. Oostra, M. et al. Topology and membrane anchoring of the coronavirus replication complex: not all hydrophobic domains of nsp3 and nsp6 are membrane spanning. *J. Virol.* **82**, 12392–12405 (2008).

16. Gautier, R., Douguet, D., Antonny, B. & Drin, G. HELIQUEST: a web server to screen sequences with specific  $\alpha$ -helical properties. *Bioinformatics* **24**, 2101–2102 (2008).
17. Lundin, A. et al. Targeting membrane-bound viral RNA synthesis reveals potent inhibition of diverse coronaviruses including the Middle East respiratory syndrome virus. *PLoS Pathog.* **10**, e1004166 (2014).
18. Hadfield, J. et al. Nextstrain: real-time tracking of pathogen evolution. *Bioinformatics* **34**, 4121–4123 (2018).
19. Twu, W. I. et al. Contribution of autophagy machinery factors to HCV and SARS-CoV-2 replication organelle formation. *Cell Rep.* **37**, 110049 (2021).
20. Tabata, K. et al. Convergent use of phosphatidic acid for hepatitis C virus and SARS-CoV-2 replication organelle formation. *Nat. Commun.* **12**, 7276 (2021).
21. Holwerda, M., V'kovski, P., Wider, M., Thiel, V. & Dijkman, R. Identification of an antiviral compound from the pandemic response box that efficiently inhibits SARS-CoV-2 infection in vitro. *Microorganisms* **8**, 1872 (2020).
22. Lee, J. Y. et al. Absolute quantitation of individual SARS-CoV-2 RNA molecules provides a new paradigm for infection dynamics and variant differences. *eLife* **11**, e74153 (2022).
23. Thorne, L. G. et al. Evolution of enhanced innate immune evasion by SARS-CoV-2. *Nature* **602**, 487–495 (2022).
24. Dias, S. S. G. et al. Lipid droplets fuel SARS-CoV-2 replication and production of inflammatory mediators. *PLoS Pathog.* **16**, e1009127 (2020).
25. Li, D. et al. The ER-localized protein DFCP1 modulates ER–Lipid droplet contact formation. *Cell Rep.* **27**, 343–358 (2019).

26. Gao, G., Sheng, Y., Yang, H., Chua, B. T. & Xu, L. DFCP1 associates with lipid droplets. *Cell Biol. Int.* **43**, 1492–1504 (2019).
27. Herker, E., Vieyres, G., Beller, M., Krahmer, N. & Bohnert, M. Lipid droplet contact sites in health and disease. *Trends Cell Biol.* **31**, 345–358 (2021).
28. Ridley, S. et al. FENS-1 and DFCP1 are FYVE domain-containing proteins with distinct functions in the endosomal and Golgi compartments. *J. Cell Sci.* **114**, 3991–4000 (2001).
29. Axe, E. L. et al. Autophagosome formation from membrane compartments enriched in phosphatidylinositol 3-phosphate and dynamically connected to the endoplasmic reticulum. *J. Cell Biol.* **182**, 685–701 (2008).
30. Rambold, A. S., Cohen, S. & Lippincott-Schwartz, J. Fatty acid trafficking in starved cells: regulation by lipid droplet lipolysis, autophagy, and mitochondrial fusion dynamics. *Dev. Cell* **32**, 678–692 (2015).
31. Marra, P. et al. The GM130 and GRASP65 Golgi proteins cycle through and define a subdomain of the intermediate compartment. *Nat. Cell Biol.* **3**, 1101–1113 (2001).
32. Jansen, M. et al. Role of ORPs in sterol transport from plasma membrane to ER and lipid droplets in mammalian cells. *Traffic* **12**, 218–231 (2011).
33. Kim, D. K. et al. A comprehensive, flexible collection of SARS-CoV-2 coding regions. *G3* **10**, 3399–3402 (2020).
34. Venditti, R. et al. The activity of Sac1 across ER–TGN contact sites requires the four-phosphate-adaptor-protein-1. *J. Cell Biol.* **218**, 783–797 (2019).
35. Bolte, S. & Cordelières, F. P. A guided tour into subcellular colocalization analysis in light microscopy. *J. Microsc.* **224**, 213–232 (2006).

36. Gilles, J. F., Dos Santos, M., Boudier, T., Bolte, S. & Heck, N. DiAna, an ImageJ tool for object-based 3D co-localization and distance analysis. *Methods* **115**, 55–64 (2017).
37. Polishchuk, E. V. & Polishchuk, R. S. Pre-embedding labeling for subcellular detection of molecules with electron microscopy. *Tissue Cell* **57**, 103–110 (2019).

## Acknowledgements

We thank P. P. M. Rede, C. B. W. G. Gripp, E. Van Anken, C. Machamer, R. De Francesco, M. Chiara, A. Ballabio, G. Diez Roux, A. Luini, R. Sitia, C. Settembre, D. Cacchiarelli, A. Grimaldi and P. Grumati for discussions and A. Iuliano for assistance with the statistical analysis. M.A.D.M. acknowledges the support of Telethon (grant TGM16CBDM13), the Italian Association for Cancer Research (grant IG2013\_14761), the European Research Council Advanced Investigator grant 670881 (SYSMET), the University of Naples Federico II (grant STAR Plus 2020 Linea 1) and the Italian Ministry of University and Research (PRIN, 2020PKLEPN). R.S.P. acknowledges the support of Telethon (grant TGM16CBDM09). R.V. acknowledges the University of Naples Federico II (grant STAR 2017 Linea 1) and the Italian Association for Cancer Research (grant MFAG 2020, code 25174). P.T.B. and T.M.L.S. acknowledge the support of the Inova program Fiocruz, Fundação de Amparo à Pesquisa do Estado do Rio de Janeiro (FAPERJ), Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) and Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES).

## Author information

### Author notes

1. These authors contributed equally: Simona Ricciardi, Andrea Maria Guarino, Laura Giaquinto

## Authors and Affiliations

1. Telethon Institute of Genetics and Medicine (TIGEM), Pozzuoli, Italy

Simona Ricciardi, Andrea Maria Guarino, Laura Giaquinto, Elena V. Polishchuk, Michele Santoro, Giuseppe Di Tullio, Cathal Wilson, Francesco Panariello, Roman S. Polishchuk, Rossella Venditti & Maria Antonietta De Matteis

2. Department of Molecular Medicine and Medical Biotechnology,  
University of Naples Federico II, Naples, Italy

Simona Ricciardi, Rossella Venditti & Maria Antonietta De Matteis

3. Laboratório de Imunofarmacologia, Instituto Oswaldo Cruz (IOC),  
Fundação Oswaldo Cruz (FIOCRUZ), Rio de Janeiro, Brazil

Vinicius C. Soares, Suelen S. G. Dias, Julia C. Santos, Thiago M. L. Souza & Patrícia T. Bozza

4. Programa de Imunologia e Inflamação, Universidade Federal do Rio de Janeiro (UFRJ), Rio de Janeiro, Brazil

Vinicius C. Soares

5. Centro de Desenvolvimento Tecnológico em Saúde (CDTS) and  
National Institute for Science and Technology on Innovation on  
Diseases of Neglected Populations (INCT/IDNP), FIOCRUZ, Rio de Janeiro, Brazil

Thiago M. L. Souza

6. Istituto Zooprofilattico Sperimentale del Mezzogiorno, Portici, Italy

Giovanna Fusco, Maurizio Viscardi & Sergio Brandi

## Contributions

M.A.D.M. conceived the work. R.V. coordinated the experimental plan.  
R.V., S.R., A.M.G. and L.G. planned and analysed most of the experiments.  
E.V.P. and R.S.P. performed the EM, CLEM and tomography analyses. M.S.

and G.D.T. developed plasmid constructs and provided technical support. A.M.G. and G.D.T. performed the protein studies. C.W. provided background data analysis. F.P. performed the VOC evolution analysis. V.C.S., S.S.G.D., J.C.S., T.M.L.S., P.T.B., G.F., M.V. and S.B. performed the SARS-CoV-2 studies. M.A.D.M. conceptualized the work and strategy and wrote the manuscript.

## Corresponding authors

Correspondence to [Roman S. Polishchuk](#), [Rossella Venditti](#) or [Maria Antonietta De Matteis](#).

## Ethics declarations

### Competing interests

The authors declare no competing interests.

## Peer review

### Peer review information

*Nature* thanks the anonymous reviewers for their contribution to the peer review of this work. [Peer reviewer reports](#) are available.

## Additional information

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

## Extended data figures and tables

[Extended Data Fig. 1 NSP6 requires a free C terminus to exert its membrane-deforming activity.](#)

**a**, Fluoromicrographs of Calu-3 cells expressing YFP-Cb5 alone (leftmost panels) or in combination with C-terminally or N-terminally Flag-tagged NSP6, or untagged NSP6 (NSP6), as indicated, and immunostained with anti-Flag antibody (red). **b**, Expression analysis of NSP6-Flag (C-term) or Flag-NSP6 (N-term) in transfected HeLa cells. Left, representative fluorescence micrographs (anti-Flag antibody). Right, fluorescence intensity measurements. Single values are plotted, Means  $\pm$  SEM are indicated.  $N = 3$ ,  $n = 69$  cells. ns, not significant. Two-tailed unpaired t-test with Welch's correction. Lower panel, western blot of total protein lysates using an anti-Flag antibody; actin was used as a loading control. NT, non-transfected cells. **c**, Western blot of HeLa cells expressing HA-NSP6 or untagged NSP6, detected using anti-HA or anti-NSP6 antibody, as indicated. NT, non-transfected cells. **d**, HeLa cells transfected with Flag-NSP6 immunostained with anti-Flag, anti-LAMP1, anti-EEA1 or anti-LC3 antibodies. **e**, Fluoromicrographs of HeLa cells expressing YFP-Cb5 and either IBV NSP6-Flag (upper panels) or IBV-Flag-NSP6 (lower panels). Cells immunostained with anti-Flag antibody (red). **f**, IEM (anti-HA immuno-gold-labelling) of a HeLa cell expressing HA-NSP6. **g**, Magnification of the boxed area. Arrows show the regular single membrane of the ER cisterna; arrowheads indicate zippered membranes in the circular NSP6-positive structure. **h**, Single slices from a tomogram of a HeLa cell expressing HA-NSP6. Connections of circular zippered structures with the ER are shown by black arrows, and the connection of linear zippered membranes with the nuclear envelope is indicated by a white arrow (see Supplementary Video 1). **i**, Routine EM of a HeLa cell expressing HA-NSP6 and **j**, magnification of boxed area. The arrow indicates apposition of limiting membranes of an ER cisterna (asterisk) that then continues into the zippered ER domain (arrowheads). Western blots in **b,c** are representative of three independent experiments each. Scale bars, **a, b, d, e**,  $10 \mu\text{m}$ ; **f, g**,  $100 \text{ nm}$ ; **h**,  $230 \text{ nm}$ ; **i**,  $200 \text{ nm}$ ; **j**,  $100 \text{ nm}$ .

### Source data

**Extended Data Fig. 2 The NSP6 compartment is accessible to ER membrane proteins with small luminal domains but not to membrane proteins with large luminal domains.**

**a**, Fluoromicrographs of HeLa cells expressing CLRT-mCherry alone (left panel) or with GFP-NSP6 (right panel). **b**, Fluoromicrographs of HeLa cells expressing GFP-ATF6 or GFP-KDEL or GFP-ERGIC53 alone (left panel) or with mCherry-NSP6 (middle and right panels). **c**, Fluoromicrographs of HeLa cells expressing GFP-ATL2 alone (left panel), or with mCherry-NSP6 (middle panel), or expressing GFP-VAP-A (right panel). **d**, HeLa cells expressing GFP-KDELR alone (left panel) or with mCherry-NSP6 (middle and right panels). Small panels in (b–d), enlargements of boxed areas, arrowheads indicate co-localization in (d). **e**, Representative images of Hela cells expressing YFP-Cb5 alone or with mCherry-NSP6, as indicated. For FRAP analysis, individual NSP6 compartments (boxed) were photobleached and the fluorescence recovery was monitored. The small panels are representative frames (from a total of 100) showing time in seconds after the bleach (see Supplementary Video 3). See Fig. 1k for FRAP measurements. **f**, Representative images from FRAP experiments of Hela cells incubated with BODIPY C<sub>12</sub>-HPC (PC), without or with mCherry-NSP6 transfection, as indicated. Individual NSP6 compartments (boxed) were photobleached and the fluorescence recovery was monitored. The small panels are representative frames (from a total of 100) showing time in seconds after the bleach (see Supplementary Videos 2 and 3). Scale bars, 10 μm.

### Extended Data Fig. 3 NSP6 undergoes homodimerization through the 1–157 region.

**a**, HeLa cells expressing C-terminal or N-terminal Flag-tagged NSP6 immunostained with anti-Flag antibody and an antibody against a luminal epitope of TGN46 after permeabilization with digitonin and subsequently with Triton-X-100 (see [Methods](#)). **b**, Model of the amphipathic helix of NSP6 (left panel) according to HELIQUEST (see [Methods](#)). Apolar residues are in yellow, polar residues and glycine have been given different colours. The arrow indicates the hydrophobic moment ( $\mu H = 0.409$ ). Numbers indicate amino acid positions of the NSP6 protein. Right panel, model of the F220Q/T222W NSP6 mutant helix ( $\mu H = 0.191$ ). Mutations that abolish the amphipathic character of the helix are in red. **c**, HeLa cells untransfected (left panel) or expressing Myc-NSP6 were immunostained for VAP-A or for Myc. Insets show the Myc-NSP6 signal. The number indicates the fraction of VAP-A associated with the NSP6 structures. Mean ± SD, N = 3, n = 74. **d**,

HeLa cells expressing GFP-NSP6 F220Q/T222W mutant. **e, f, g**, Cell lysates (input) and immunoprecipitates (IP, with anti-HA or anti-Flag antibodies) from HeLa cells, untransfected or expressing the indicated NSPs were analysed by western blot with anti-HA, anti-Flag or anti-GFP antibodies as appropriate. Images are representative of three independent experiments. **h**, Fluoromicrographs of HeLa cells expressing GFP-NSP6(1–157) alone or with mCherry-NSP6. Scale bar, **a, c, d, h**, 10 μm.

### Extended Data Fig. 4 K22 interferes with the formation of the NSP6 compartment.

**a**, Stably transfected Flag-NSP6 clone induced with doxycycline and treated with DMSO or K22 for 24 h. K22 reduced the number of NSP6 structures and resulted in elongated structures in a percentage of the cells (right panel, number of cells exhibiting these structures. Mean ± SD). The number of NSP6 structures in DMSO and K22-treated cells is plotted as single values. Mean ± SEM, N = 3, n = 90. Two-tailed unpaired t-test with Welch's correction. **b**, FRAP analysis of mCherry-NSP6-structures (boxed) in cells treated with DMSO or K22. The small panels show time in seconds after the bleach. Graph, FRAP curves expressed as a % of time 0. Mean ± SD, N = 3, n = 45 structures. **c**, Cells expressing GFP-VAP-A alone (as a control) or with Flag-NSP6, treated with DMSO or with K22 for 16 h. **d–f**, Immuno-CLEM of HA-NSP6-expressing cells treated with K22 for 24 h. **d**, Fluoromicrograph showing NSP6 (anti-HA immunostaining with Alexa Fluor®546-FluoroNanogold secondary antibody) in elongated structures (arrows 1, 2) close to the nucleus (asterisk). **e**, EM section of the same cells (asterisks) shown in (**d**). Arrows 1 and 2 indicate overlap of the fluorescent and immuno-gold signals in the elongated zippered domains of the nuclear envelope (NE). **f**, Serial sections of the structure indicated by arrow 1 in panel (**e**). White arrows: NSP6-zippered domains of the NE, black arrows: regular NE. Insets, magnification of boxed areas showing regular NE (arrowheads: nuclear pore). **g, h**, EM showing regular (black arrows) and zippered (white arrows) NE domains in cells expressing HA-NSP6 treated with DMSO (**g**) or K22 (**h**). **i**, A cell not expressing NSP6 treated with K22 shows regular NE (black arrows). Insets in **g–i**, magnification of boxed areas showing regular NE (arrowheads: nuclear pores). **j**, Morphometric analysis of zippered NE surface in control and K22-treated cells expressing HA-

NSP6. Single values are plotted, Medians are indicated, n = 20 cells, two-tailed unpaired t-test. Scale bars, **a–c**, 10  $\mu$ m; **d**, 7.5  $\mu$ m; **e**, 3.8  $\mu$ m; **f**, 750 nm, inset 200 nm; **g–i**, 1  $\mu$ m, insets 200 nm.

[Source data](#)

**Extended Data Fig. 5 NSP(ΔSGF) is more prone to homodimerization and/or oligomerization than the reference NSP6.**

**a**, Fluoromicrographs of stably expressing Flag-NSP6 or Flag-NSP6(ΔSGF) cells induced with doxycycline at the indicated times. **b**, Levels of HA-NSP6 and HA-NSP6(ΔSGF) clones induced overnight with doxycycline analysed by western blot with anti-HA antibody. GAPDH was used as loading control. **c**, Doxycycline-induced HeLa clones expressing HA-NSP6 or HA-NSP6(ΔSGF), or parental (CTRL) cells, were radiolabelled for 1 h with  $^{35}$ S-methionine/cysteine and chased for the indicated times. Samples were immunoprecipitated and analysed by SDS-PAGE gel autoradiography (top panels) and by western blot (bottom panels). The estimated half-life for HA-NSP6 and HA-NSP6(ΔSGF) was 5 h. **d**, Western blot of supernatant (S) and pellet (P) of lysates (at increasing Triton-X-100/NP-40 concentrations) of cells expressing NSP6-Flag, Flag-NSP6, or Flag-NSP6(ΔSGF) (see [Methods](#)). Numbers indicate the fraction of protein in the supernatant. **e**, Cell lysates (input) of cells expressing GFP-NSP6 with HA-NSP6, or GFP-NSP6(ΔSGF) with HA-NSP6(ΔSGF), were immunoprecipitated (IP) with anti-HA antibody and analysed by western blot with anti-HA and anti-GFP antibodies. The graph shows co-IP efficiency of NSP6(ΔSGF) relative to NSP6, which was set as 1 (see [Methods](#)). Mean  $\pm$  SEM, n = 3 samples examined over three independent experiments. Two-tailed unpaired t-test with Welch's correction. **f**, Fluorescence Loss in Photobleaching (FLIP) analysis of cells expressing GFP-NSP6 or GFP-NSP6(ΔSGF). Left panels, before bleaching. Right panels, after bleaching. Dashed lines indicate the areas in which iterative bleaching was applied. Graph: quantitative analysis of FLIP (see [Methods](#)). Values are expressed as a percentage of time 0, Means  $\pm$  SD, three independent experiments, n = 10–12 cells per experiment. The calculated FLIP half-life for GFP-NSP6 is 70.5 s  $\pm$  12.6 and for GFP-NSP6(ΔSGF) 103.8 s  $\pm$  27.6. **g**, FRAP analysis of GFP-

NSP6( $\Delta$ SGF)-expressing cells treated with DMSO or K22 for 16 h after bleaching of the individual NSP6( $\Delta$ SGF)-compartments (boxed). The small panels are representative frames (from a total of 100) at different times (in seconds) after the bleach. **h**, Quantitative FRAP analysis of the experiment in (g). Fluorescence intensity is expressed as a percentage of the value measured at time 0 (normalized to 100%). Means  $\pm$  SD, three independent experiments, n = 45 structures. **i**, Doxycycline-induced clone expressing Flag-NSP6( $\Delta$ SGF) treated with DMSO or K22 for 24 h. Number of NSP6 structures (middle panel) and cells with elongated NSP6 structures (right panel) induced by K22. The number indicates the percentage (Mean  $\pm$  SD) of cells exhibiting the elongated structures. Single values are plotted, Means  $\pm$  SEM are indicated, N = 3, n = 90, two-tailed unpaired t-test with Welch's correction. **j–l**. Immuno-CLEM of cells expressing HA-NSP6( $\Delta$ SGF). **j**, Fluoromicrograph (anti-HA immunostaining) of the NSP6 compartment. **k**, Magnification of the box in (j) (inset) and IEM, in which arrows 1–4 indicate overlap of the fluorescent and immuno-gold signals in the zippered NSP6-positive-structures. **l**, magnification of the structure indicated by arrow 1. **m**, Fluoromicrographs of cells expressing Flag-NSP6/NSP7-V5 or Flag-NSP6( $\Delta$ SGF)/NSP7-V5 immunostained with anti-Flag (green), anti-V5 (white insets) and anti-GM130 antibodies (red insets). Bottom panels, merge of Flag-NSP6 and GM130. **n**, Western blot of cell lysates from non-transfected (NT) and HA-NSP6/NSP7-V5 or HA-NSP6( $\Delta$ SGF)/NSP7-V5 expressing HeLa cells. Western blot images are representative of three independent experiments. Scale bars, **a**, **f**, **g**, **i**, **m**, 10  $\mu$ m; **j**, 3.7  $\mu$ m; **k**, 480 nm; **l**, 250 nm.

### [Source data](#)

## [Extended Data Fig. 6 NSP6 organizes the DMVs induced by NSP3–NSP4.](#)

**a**, Fluoromicrographs of HeLa cells expressing YFP-Cb5 with HA-NSP3 (anti-HA immunostaining) or mCherry-NSP4. Insets, merge with YFP-Cb5. **b**, Fluoromicrographs of HeLa cells expressing HA-NSP3 and mCherry-NSP4. Insets, enlargement of boxed area. Arrowheads, NSP3/NSP4-positive structures. Dashed lines delineate cell boundaries. **c**, Western blot (WB) of total lysates from HeLa cells expressing HA-NSP3, mCherry-NSP4, Flag-

NSP6, or GFP as indicated. Actin was used as loading control. **d**, IEM and **e**, routine EM of HeLa cells co-transfected with HA-NSP3 and mCherry-NSP4. Anti-HA labelling in (**d**) shows gold particles decorating DMVs, indicated by asterisks. Black arrows, ER. Inset, magnification of boxed area. White arrows in **d** and **e** show double membranes. The average DMV size is  $92 \pm 30$  nm. **f**, Western blot of total lysates from HeLa cells expressing HA-NSP3, mCherry-NSP4 or Flag-NSP6 as indicated. Actin was used as loading control. **g**, Individual fluoromicrographs of a Calu-3 cell co-transfected with Flag-NSP6, HA-NSP3, and mCherry-NSP4. **h**, Length of DMV–ER tubular or zippered connections in NSP3/NSP4 or in NSP3/NSP4/NSP6-expressing cells, respectively. Single values are plotted. Medians are shown ( $n \geq 14$  connections), two-tailed unpaired t-test. **i**, Tomographic slice of a HeLa cell expressing HA-NSP3/mCherry-NSP4 or **j**, HA-NSP3/mCherry-NSP4/Flag-NSP6, showing DMV clusters with regular round DMVs (white arrows) and large and elongated DMVs (black arrows). **k**, Frequency histograms of DMV diameter measured from tomograms of cells expressing NSP3/NSP4 (average diameter 80.87 nm) or NSP3/NSP4/NSP6 (average diameter 67.50 nm). Non-parametric Kolmogorov-Smirnov (KS) test.  $n \geq 135$  vesicles. **l**, Tomographic slice of HeLa cells transfected with HA-NSP3/mCherry-NSP4 or **m**, HA-NSP3/mCherry-NSP4/Flag-NSP6, with arrows indicating the edges of the DMV clusters. **n**, DMV densities were calculated in tomograms as the number of vesicles per  $\mu\text{m}^3$  in a volume occupied by a DMV cluster.  $n = 8$  clusters; Single values are plotted, Medians are shown, two-tailed unpaired t-test. Scale bars, **a**, **b**, **g**,  $10 \mu\text{m}$ ; **i**, **j**, **l**, **m**,  $180$  nm.

### Source data

### Extended Data Fig. 7 K22 impairs the ability of NSP6 to organize the NSP3–NSP4 puncta and has anti-SARS-CoV-2 activity.

**a**, HeLa cells transfected with HA-NSP3 and mCherry-NSP4 for 5 h were further transfected or not with Flag-NSP6 and treated with DMSO or K22 (40  $\mu\text{M}$ , 16 h) followed by immunostaining as indicated. **b**, Quantification of the number of NSP4 puncta/cell in (**a**).  $N = 3$ ,  $n = 60$ . Single values are plotted. The median value is shown. One-way ANOVA test with Tukey's post-hoc. ns, not significant. **c**, **d**, CLEM analysis of K22-treated cells. **c**,

Fluoromicrograph of HeLa cell expressing HA-NSP3, mCherry-NSP4 and Myc-NSP6. Inset corresponds to the boxed area and shows NSP3/NSP4-positive structures (arrows) close to the NSP6 compartment (arrowhead). **d**, Overlap of fluorescent image (inset in **c**) with EM image. The NSP6 compartment corresponds to a circular zippered ER structure (arrowhead) close to but not connected with the NSP3/NSP4 puncta that correspond to DMVs (arrows). The empty arrow indicates a tubular connection of a DMV to the regular ER (magnified in the inset). **e**, Ultrastructure of DMV clusters in K22-treated cells expressing HA-NSP3, mCherry-NSP4 and Myc-NSP6. Serial sections show a DMV cluster with irregular elongated DMVs (black arrows). The empty arrow indicates a tubular connection of a DMV with regular ER. **f**, Morphometric analysis of serial sections from untreated (NT) and K22-treated cells to quantify the number of DMVs per cluster, DMV shape factor, and the number of tubular or zippered connections per DMV cluster. Single values are plotted, Medians are shown, n = at least 8 clusters or 70 DMVs, two-tailed unpaired t-test. **g, h**, Antiviral activity of K22. Effects of K22 on cell death measured by LDH (**g**) or on viral replication (**h**) in SARS-CoV-2-infected Calu-3 cells. Mean  $\pm$  SEM, N = 3, one-way ANOVA. **i**, Calu-3 cells infected with SARS-CoV-2 without (left panels) or with (right panel) K22 treatment. Cells were immunostained for dsRNA and NSP6. Nuclei were stained with DAPI. An enlargement of the boxed area shows NSP6 labelling (arrowheads) in the proximity of replication areas labelled by dsRNA. Graph, NSP6 structures within a distance of 250 nm from dsRNA spots, and dsRNA spots within a distance of 250 nm from NSP6 structures were counted and expressed as a percentage of total. Mean  $\pm$  SEM, 15 cells analysed, n = 729 NSP6 structures, n = 901 dsRNA spots. Scale bar, **a**, 10  $\mu$ m; **c**, 4.4  $\mu$ m; **d**, 370 nm; **e**, 320 nm; **i**, 20  $\mu$ m.

#### [Source data](#)

#### [Extended Data Fig. 8 NSP6\( \$\Delta\$ SGF\) is more proficient in zippering the ER.](#)

**a**, Tomographic slice of a HeLa cell expressing HA-NSP3/mCherry-NSP4/Flag-NSP6 or **b**, expressing HA-NSP3/mCherry-NSP4/Flag-NSP6( $\Delta$ SGF). Arrows indicate zippered ER connectors directed towards DMV clusters. **c**, Quantification of the number of ER zippered connections

per DMV cluster and **d**, number of DMVs per cluster. Single values are plotted, Medians are shown. NSP3/NSP4/NSP6, N = 7 cells, n = 8 DMV clusters; NSP3/NSP4/NSP6( $\Delta$ SGF), N = 8 cells, n = 9 DMV clusters. Two-tailed unpaired t-test. **e**, Frequency histograms of DMV diameter measured from tomograms of cells expressing NSP3/NSP4/NSP6 (average diameter 67.50 nm) or NSP3/NSP4/NSP6( $\Delta$ SGF) (average diameter 68.51 nm). The histograms were analysed using non-parametric Kolmogorov-Smirnov (KS) test. n  $\geq$  123 vesicles. **f**, Serial tomographic slices from Calu-3 cells infected with 10 MOIs of an early lineage SARS-CoV-2 for 24 h. White arrows indicate a zippered connector that links ER (black arrows) to two DMVs (1 and 2). **g**, **h**, Serial tomographic slices from Calu-3 cells infected with the Gamma variant of SARS-CoV-2. White arrows indicate zippered connectors that depart from the ER (black arrows) and then branch (red arrow) towards two DMVs (1 and 2) in **(g)** and link the ER to a DMV (1) in **h**. White arrowheads indicate connectors that link DMV2 to DMV3 in **(g)** and DMV1 to DMV3 in **(h)**. Scale bars, **a**, **b** 140 nm; **f–h** 200 nm.

#### [Source data](#)

#### [\*\*Extended Data Fig. 9 RAB18 is recruited to ROLs by NSP6.\*\*](#)

**a**, Fluoromicrographs of HeLa cells expressing Flag-NSP6-C80 (i.e. the last 80 amino acids of NSP6). Anti-Flag antibody and staining for LDs using BODIPY C<sub>12</sub>. Insets, enlargement of boxed areas. Arrowheads, colocalizing structures. **b**, Flag-NSP6-C80 associates with roundish structures that stain with BODIPY-488 and with ADRP (perilipin 2). **c**, NSP6-C80 mutated in residues that abrogate the amphiphilic properties of the AH fail to associate with LDs. Cells were immunostained with anti-Flag Ab (green) and with BODIPY-DA-C<sub>12</sub> (red). **d**, Calu-3 cells infected with SARS-CoV-2 stained for LDs (Oil Red O staining)<sup>22</sup> and immunostained with anti-dsRNA and anti-NSP6 antibodies. Blue, nuclear DAPI staining. Graph, quantification of the association of NSP6 or dsRNA positive structures with LDs, expressed as a percentage of the total number of NSP6 or dsRNA structures per cell. Mean  $\pm$  SEM, N = 12, n = 1,377 and n = 861 for dsRNA and NSP6, respectively. **e**, Calu-3 cells infected or not with SARS-CoV-2, with or without the DGAT-1 inhibitor A922500, were analysed for LDs using Oil Red O staining (measurement of the fluorescent area of LDs- left graph) or

for viral titres<sup>22</sup> (right graph). Two-tailed unpaired t-test with Welch's correction (left graph); one-way ANOVA followed by Tukey's post-hoc test (right graph). **f**, Graph obtained applying the function Shuffle (see [Methods](#)) for the “NSP6 objects” and “LD objects” in the NSP3/NSP4/NSP6 transfected cell in Fig. [4a](#), showing that the measured distances (blue line) are significantly different from mean random distances (red line) flanked by 95% confidence intervals (green lines). **g**, Fluoromicrographs of Calu-3 cells expressing mCherry-NSP4, HA-NSP3, and Flag-NSP6 (left panel), or mCherry-NSP4 and HA-NSP3 (middle panel), or Flag-NSP6 (right panel). Cells were immunostained with anti-Flag Ab (blue). mCherry fluorescence was used as a read-out for mCherry-NSP4/HA-NSP3 structures. LDs were detected using BODIPY-488 (green). Inset, enlargement of boxed area. Arrowheads indicate LDs close to the NSP6 compartment and in proximity to the replication organelle. **h**, Cells expressing GFP-RAB18 alone or with Flag-NSP6 (anti-Flag immunostaining). Insets, arrowheads show co-localization of RAB18 with the NSP6 compartment. **i**, Individual fluoromicrographs and merge of cells co-expressing GFP-DFCP1, mCherry-NSP4 and HA-NSP3 (anti-HA immunostaining). **j**, In vitro pull-down assays using total cell lysates (input) from HeLa cells expressing HA-NSP6 incubated with GST-DFCP1 or GST alone. Upper panel, Ponceau Red staining; bottom panel, western blot with anti-HA antibody. Images are representative of three independent experiments. Scale bars, **a–c, g–i**, 10 μm; **d**, 20 μm.

#### [Source data](#)

### [Extended Data Fig. 10 The C-terminal domain of NSP6 is involved in the recruitment of DFCP1 in a PtdIns3P-independent manner.](#)

**a**, Fluoromicrographs of cells co-expressing GFP-DFCP1 and Flag-NSP6(1–157). **b**, Fluorescent images of HeLa cells transfected with the indicated DFCP1 mutants alone or in combination with NSP6 (as indicated). A schematic representation of DFCP1 mutants is reported on top. Arrowheads, DFCP1 signal in the NSP6 compartment. Insets, enlarged merge of boxed areas. Numbers indicate the percentage of co-localization between DFCP1 mutants and NSP6. Mean ± SD (see [Methods](#)). **c**, Fluorescent micrographs

of cells expressing mCherry-DFCP1 and Flag-NSP6 treated with SAR405 or with wortmannin. Anti-Flag immunostaining. **d**, Fluoromicrograph showing LC3 staining in one transfected and one non-transfected cell from Flag-NSP6-expressing HeLa cells. Inset, anti-Flag immunostaining. Graph, quantification of LC3 puncta in non-transfected (CTRL) and NSP6-transfected cells (number of LC3 spots/cell). Means  $\pm$  SEM. n = 74 cells examined over three independent experiments. Two-tailed unpaired t-test with Welch's correction. ns, not significant. **e**, HeLa cells expressing HA-NSP3 and GFP-NSP4, or HA-NSP3, GFP-NSP4 and Flag-NSP6, were loaded with BODIPY-DA-C<sub>12</sub> and washed out for 6 h (see [Methods](#)). Dotted yellow lines, LDs; dotted white lines, NSP4 puncta. Graph, quantification of the percentage of NSP4 puncta positive for BODIPY-DA-C<sub>12</sub>. Single values are plotted, Means  $\pm$  SEM are indicated, N = 3, n = 90, two-tailed Mann-Whitney test. **f**, Western blot of protein extracts from mock-treated (CTRL) and DFCP1-KD cells. Actin was used as loading control. The graph shows the level of DFCP1 protein expressed as percentage of control (set at 100). Mean  $\pm$  SEM. N = 3, Two-tailed unpaired t-test with Welch's correction. **g**, Western blot of protein lysates from untreated (CTRL), scramble- and DFCP1 siRNA-treated (DFCP1-KD) SARS-CoV-2-infected Calu-3 cells detected with an anti-DFCP1 antibody. GAPDH was used as loading control. The graph shows the level of DFCP1-KD expressed as percentage of control (set at 100). Mean  $\pm$  SEM. N = 3, two-tailed unpaired t-test with Welch's correction. **h**, Quantification of viral titres in SARS-CoV-2-infected Calu-3 cells untreated (-), transfected with scramble siRNA (CTRL) or DFCP1 siRNA (DFCP1-KD). Mean  $\pm$  SEM, N = 7, two-tailed unpaired t-test with Welch's correction. Scale bar **a–e**, 10  $\mu$ m. Western blot images are representative of three independent experiments.

#### [Source data](#)

#### [Extended Data Fig. 11 Working model for the role of NSP6 in replication organelle biogenesis.](#)

NSP6-induced zippered connectors are cues and organizers for NSP3/NSP4-induced DMV formation acting as selective communication tracks with the ER (largely excluding luminal ER proteins). In addition, the connectors

might also act as fast-tracks to refurbish the actively growing subpopulation of DMVs with lipids derived from LDs.

## Supplementary information

### Supplementary Information

This file contains legends for Supplementary Videos 1–14, Supplementary Tables 1–2 and Supplementary Figure 1.

### Reporting Summary

### Peer Review File

### Video 1 Electron tomogram of HeLa cells transfected with HA–NSP6.

Connections of circular zippered structures with the ER are shown by black arrows, and the connection of linear zippered membranes with the nuclear envelope is indicated by a white arrow.

### Video 2 FRAP experiment of HeLa cells transfected with GFP–VAP-A and mCherry–NSP6.

The bleached region is highlighted in the boxed area.

### Video 3 FRAP experiment of HeLa cells transfected with YFP–Cb5 and mCherry–NSP6.

The bleached region is highlighted in the boxed area.

### Video 4 FRAP experiment of HeLa cells transfected with mCherry–NSP6 and incubated with BODIPY C<sub>12</sub>-HPC.

The bleached region is highlighted in the boxed area.

## **Video 5 Electron tomogram of HeLa cells transfected with HA–NSP3, mCherry–NSP4, and GFP–NSP6.**

White arrows show connections of zippered ER to DVMs. Black arrows show connection of zippered region to regular ER cisternae.

## **Video 6 3D reconstruction of the tomogram shown in EM Video 5.**

DMVs, zippered ER, and regular ER are shown in red, green and yellow, respectively.

## **Video 7 Electron tomogram of HeLa cells transfected with HA–NSP3, mCherry–NSP4, and GFP–NSP6.**

Another example from electron tomogram of connections between zippered ER and DVMs (white arrows) and between zippered ER and regular ER cisternae (black arrow).

## **Video 8 Electron tomogram of HeLa cells transfected with HA–NSP3 and mCherry–NSP4.**

White arrows show connections between ER and DVMs.

## **Video 9 3D reconstruction of the tomogram shown in EM Video 8.**

DMVs and ER are shown in red and yellow, respectively.

## **Video 10 Electron tomogram of HeLa cells transfected with HA–NSP3 and mCherry–NSP4.**

Another example from electron tomogram of connections between ER and DVMs (white arrows).

## [Video 11 Electron tomogram of HeLa cells transfected with HA–NSP3, mCherry–NSP4, and GFP–NSP6\(ΔSGF\).](#)

White arrows show numerous connections of zippered ER to DVMs. Black arrow shows connection of zippered region to regular ER cisternae.

## [Video 12 3D reconstruction of the tomogram shown in EM Video 11.](#)

DMVs, zippered ER, and regular ER are shown in red, green and yellow, respectively.

## [Video 13 Electron tomogram of Calu-3 cells infected with early lineage B.1 SARS-CoV-2.](#)

White arrow shows zippered DMV connector.

## [Video 14 Electron tomogram of Calu-3 cells infected with the Gamma variant of SARS-CoV-2.](#)

White arrows show long zippered ER connectors directed to DMVs.

## **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. 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

[Reprints and Permissions](#)

## About this article

### Cite this article

Ricciardi, S., Guarino, A.M., Giaquinto, L. *et al.* The role of NSP6 in the biogenesis of the SARS-CoV-2 replication organelle. *Nature* **606**, 761–768 (2022). <https://doi.org/10.1038/s41586-022-04835-6>

- Received: 12 April 2021
- Accepted: 05 May 2022
- Published: 12 May 2022
- Issue Date: 23 June 2022
- DOI: <https://doi.org/10.1038/s41586-022-04835-6>

## Share this article

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

[Get shareable link](#)

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

[Copy to clipboard](#)

Provided by the Springer Nature SharedIt content-sharing initiative

---

This article was downloaded by **calibre** from <https://www.nature.com/articles/s41586-022-04835-6>

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

- Article
- Open Access
- [Published: 08 June 2022](#)

# Cohesin-mediated loop anchors confine the locations of human replication origins

- [Daniel J. Emerson](#),
- [Peiyao A. Zhao](#),
- [Ashley L. Cook](#),
- [R. Jordan Barnett](#),
- [Kyle N. Klein](#),
- [Dalila Saulebekova](#),
- [Chunmin Ge](#),
- [Linda Zhou](#),
- [Zoltan Simandi](#),
- [Miriam K. Minsk](#),
- [Katelyn R. Titus](#),
- [Weitao Wang](#),
- [Wanfeng Gong](#),
- [Di Zhang](#),
- [Liyan Yang](#),
- [Sergey V. Veney](#),
- [Johan H. Gibcus](#),
- [Hongbo Yang](#),
- [Takayo Sasaki](#),
- [Masato T. Kanemaki](#),
- [Feng Yue](#),
- [Job Dekker](#),
- [Chun-Long Chen](#),
- [David M. Gilbert](#) &

- [Jennifer E. Phillips-Cremins](#)

[Nature](#) volume **606**, pages 812–819 (2022)

- 7263 Accesses
- 73 Altmetric
- [Metrics details](#)

## Abstract

DNA replication occurs through an intricately regulated series of molecular events and is fundamental for genome stability<sup>1,2</sup>. At present, it is unknown how the locations of replication origins are determined in the human genome. Here we dissect the role of topologically associating domains (TADs)<sup>3,4,5,6</sup>, subTADs<sup>7</sup> and loops<sup>8</sup> in the positioning of replication initiation zones (IZs). We stratify TADs and subTADs by the presence of corner-dots indicative of loops and the orientation of CTCF motifs. We find that high-efficiency, early replicating IZs localize to boundaries between adjacent corner-dot TADs anchored by high-density arrays of divergently and convergently oriented CTCF motifs. By contrast, low-efficiency IZs localize to weaker dotless boundaries. Following ablation of cohesin-mediated loop extrusion during G1, high-efficiency IZs become diffuse and delocalized at boundaries with complex CTCF motif orientations. Moreover, G1 knockdown of the cohesin unloading factor WAPL results in gained long-range loops and narrowed localization of IZs at the same boundaries. Finally, targeted deletion or insertion of specific boundaries causes local replication timing shifts consistent with IZ loss or gain, respectively. Our data support a model in which cohesin-mediated loop extrusion and stalling at a subset of genetically encoded TAD and subTAD boundaries is an essential determinant of the locations of replication origins in human S phase.

## Main

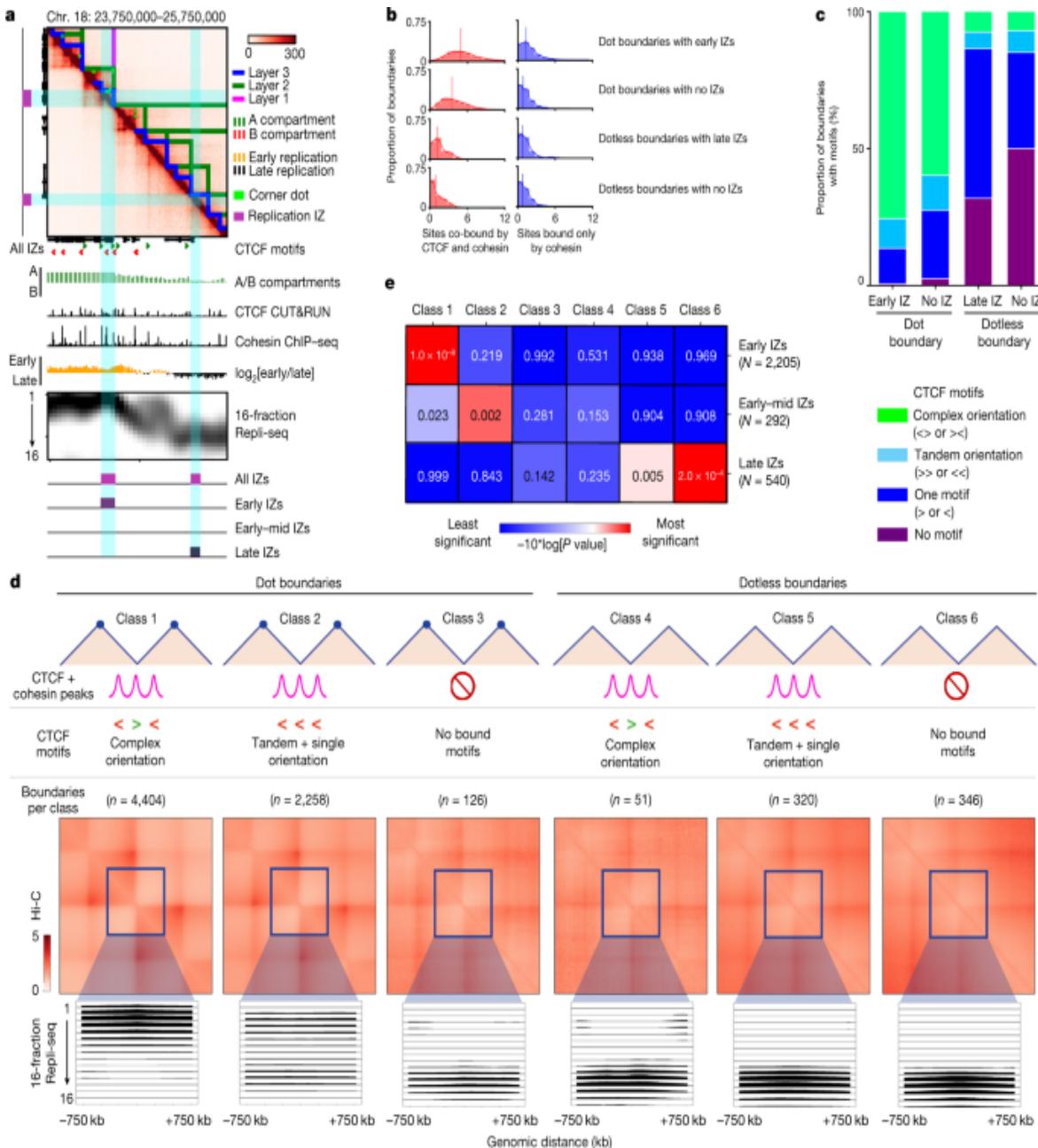
The interphase human genome folds into TADs and nested subTADs. TADs were originally defined in first-generation Hi-C and 5C data as megabase (Mb)-scale, self-interacting chromatin segments in which DNA sequences exhibit substantially higher contact frequency within—compared to between—domains<sup>3,4,5,6</sup>. Molecular and computational advances over the past decade have resulted in ultrahigh-resolution genome folding maps with substantially improved signal-to-noise ratios<sup>8,9,10,11</sup>. Such technical advances have enabled the discovery of fine-grained A/B compartments<sup>8</sup>, nested subTADs within TADs<sup>7</sup>, punctate dot structures indicative of long-range looping interactions<sup>8</sup>, and stripes indicative of loop extrusion<sup>12,13,14</sup>. In light of the critical importance of dissecting the link between specific higher-order chromatin architectural features and genome function, a leading challenge is to classify subtypes of TADs/subTADs in Hi-C maps by their fine-grained structural features. Clearly defining structural classes of TADs/subTADs can in turn facilitate the careful dissection of each boundary’s molecular composition, organizing principles and unique cause-and-effect relationship across a range of genome functions.

Here we ascertain the functional link between distinct structural classes of TADs/subTADs and DNA replication. Replication initiates from tens of thousands of origins licensed in excess across the human genome in telophase and throughout G1 (refs.<sup>1,2</sup>). A small proportion of licensed origins subsequently fire in orchestrated temporal waves during S phase<sup>2</sup>. It is established that origins fire at one or more sites chosen stochastically within ≈40 kb regions (IZs)<sup>15,16,17</sup>. Nevertheless, a consensus sequence encoding origin or IZ placement has not been definitively identified in humans. Waves of early and late replication correlate with A and B compartments, respectively, and the temporal transitions from early to late replication can in some cases align with TAD boundaries<sup>3,18,19</sup>. However, the role of fine-scale genome folding patterns during interphase (such as loops, subTADs and TADs detectable in high-resolution Hi-C data) in the genomic placement of initiated origins following entry into S phase is not known.

We recently developed a high-resolution Repli-seq method to identify the placement of IZs across the genome at 50-kb resolution<sup>16</sup>. We first compared the genomic locations of IZs replicating across early, early–mid and late S

phase to our high-resolution Hi-C data developed in the 4D Nucleome Consortium from H1 human embryonic stem (ES) cells<sup>11</sup>. We noticed that high-efficiency, early-S-phase IZs colocalize to strongly insulated boundaries demarcated by corner-dot TADs/subTADs on one or both sides (Fig. [1a](#) and Extended Data Figs. [1](#), [2a](#) and [3](#)). By contrast, low-efficiency IZs that fire late in S phase can colocalize with boundaries between TADs/subTADs devoid of corner-dots (Fig. [1a](#) and Extended Data Figs. [2b](#) and [3](#)). Our qualitative observations suggest that early and late IZs are enriched at genomic locations serving as boundaries of corner-dot and dotless TAD/subTADs, respectively.

**Fig. 1: High-efficiency IZs localize specifically to corner-dot TAD/subTAD boundaries with high-density arrays of CTCF + cohesin-binding sites in complex orientations.**



**a**, A Hi-C map from H1 human ES cells for the locus chromosome (chr.) 18: 23.75 Mb–25.75 Mb, hg38, showing TADs, subTADs, loops, CTCF motifs, A/B compartments, CTCF cleavage under targets and release using nuclease (CUT&RUN), cohesin chromatin immunoprecipitation with sequencing (ChIP-seq), two-fraction Repli-seq, 16-fraction Repli-seq and IZs. **b**, Distribution of the number of sites co-bound by CTCF and cohesin (red) or bound only by cohesin (blue) per boundary for: dot boundaries colocalized with early IZs ( $n = 2,200$ ); dot boundaries colocalized with no IZs ( $n =$

4,087); dotless boundaries colocalized with late IZs ( $n = 66$ ); and dotless boundaries colocalized with no IZs ( $n = 628$ ). **c**, Proportion of boundaries with no CTCF motif, one single CTCF motif, CTCF motifs in a tandem orientation, and CTCF motifs in a complex divergent or convergent orientation. Boundaries are stratified into dot and dotless boundaries with either early/late IZs or no IZs. **d**, Top: boundary classification as detailed in the [Supplementary Methods](#) and Supplementary Table 7. Middle: aggregate peak analysis of the average observed/expected interaction frequency of the domains centred on each boundary classification. Bottom: averaged 16-fraction Repli-seq signal for each S-phase fraction centred on boundaries  $\pm 750$  kb. Boundaries, TADs and Repli-seq data were normalized to the same genomic length scale. Boundary numbers are provided only for autosomal chromosomes. **e**, We computed right-tailed, one-tailed empirical  $P$  values using a randomization test with early-, early–mid- and late-S-phase IZs and size- and A/B compartment-matched null IZs ([Supplementary Methods](#)).

To quantify the link between TAD/subTAD boundaries and IZ genomic placement, we identified a total of 23,851 chromatin domains genome-wide in Hi-C data for human ES cells using our graph-theory-based method 3DNetMod<sup>20</sup> ([Supplementary Methods](#) and Supplementary Table 1). We also applied statistical methods developed by our laboratory and others to identify dot-like structures representative of bona fide looping interactions<sup>8,21,22</sup>. We identified 16,922 dots genome wide in ensemble Hi-C maps of human ES cells. Such dots represent punctate groups of adjacent pixels with significantly higher contact frequency compared to the surrounding local chromatin domain structure (Fig. 1a, green rectangles, [Supplementary Methods](#) and Supplementary Table 2). After co-registration of dots with domains, we identified 8,279 corner-dot TADs/subTADs and 15,572 dotless TADs/subTADs genome wide in human ES cells (Supplementary Table 3). We stratified boundaries into three groups, including those that are structurally demarcated by: adjacent corner-dot TADs/subTADs on both sides (double-dot boundaries,  $n = 6,318$ ); corner-dot TADs/subTADs on only one side and dotless on the other (single-dot boundaries,  $n = 2,163$ ); and adjacent dotless TADs/subTADs on both sides (dotless boundaries,  $n = 1,089$ ) (Supplementary Table 4). By applying a range of parameter stringencies and methods for dot calling, we could modify the proportion of boundaries classified as double-dot, single-dot and

dotless, but the colocalization of dot boundaries with IZs was evident regardless of statistical methodology ([Supplementary Methods](#) and Extended Data Fig. 4). We combined all double-dot and single-dot boundaries into dot boundaries, as they showed similar IZ localization patterns (Supplementary Table 4).

Cohesin is essential for the formation of TADs/subTADs through loop extrusion and stalling against boundaries insulated by the architectural protein CTCF<sup>12,13,23,24,25</sup>. We reasoned that the density and orientation of CTCF-binding sites might reveal an architectural protein signature at boundaries linked to placement of active origins that fire in S phase. We observed a substantially higher density of co-bound CTCF + cohesin-binding sites at dot boundaries overlapping early IZs compared to those that do not overlap any IZs (Fig. 1b and Supplementary Tables 5 and 6). We also examined sites that bind only cohesin, as they can earmark CTCF-independent enhancer–promoter interactions<sup>7,23</sup>, but we did not see a notable difference in the number of sites that bind only cohesin across dot versus dotless TAD/subTAD boundaries (Fig. 1b). Together, our data indicate that boundaries colocalizing with human early-S-phase IZs exhibit enriched occupancy of motifs co-bound by CTCF and cohesin, but not cohesin alone, thus confirming and substantially expanding on observations in previous reports linking cohesin generally to a small subset of replication origins in *Drosophila*<sup>26</sup> and humans<sup>27</sup>.

Recent reports have uncovered that convergently oriented CTCF motifs anchor long-range looping interactions formed by cohesin-mediated extrusion<sup>12,14,23,28,29</sup>. We observed that most dot boundaries are marked by two or more CTCF + cohesin-bound motifs arranged in a convergent or divergent orientation (hereafter called complex motif orientation; Fig. 1c), and this molecular signature was further enriched when dot boundaries colocalize with early replicating IZs. By contrast, nearly all dotless boundaries have only one or no CTCF + cohesin-bound motifs (Fig. 1c). Dotless boundaries colocalized with late IZs were most often anchored by one CTCF motif. We therefore establish six boundary classes by stratifying dot (classes 1–3) and dotless (classes 4–6) boundaries into those localized with CTCF + cohesin-bound motifs in a complex orientation (classes 1 and

4), tandem or single-motif orientation (classes 2 and 5), or no bound motifs (classes 3 and 6; Fig. [1d](#)).

We next formulated a statistical test to quantify IZ enrichment at boundaries compared to the background expectation across autosomes ([Supplementary Methods](#) and Supplementary Table [7](#)). Consistent with our qualitative observations, high-efficiency IZs firing in early S phase were significantly enriched at dot boundaries marked by CTCF + cohesin-binding sites in complex orientations compared to a null distribution of random intervals matched by size and A/B compartment distribution (class 1; Fig. [1d,e](#), Extended Data Fig. [5b-d](#) and [Supplementary Methods](#)). By contrast, low-efficiency IZs firing in late S phase were depleted at dot boundaries and significantly enriched at dotless boundaries with tandem + single CTCF + cohesin-bound motifs or no bound motifs (classes 5 and 6; Fig. [1d,e](#), Extended Data Fig. [5b-d](#) and [Supplementary Methods](#)). We note that our null distribution was created with random intervals matched to real IZs by their size and compartment distribution, reinforcing that the enrichment reflects a strong localization at boundaries above the known link between early and late replication and A and B compartments, respectively ([Supplementary Methods](#)).

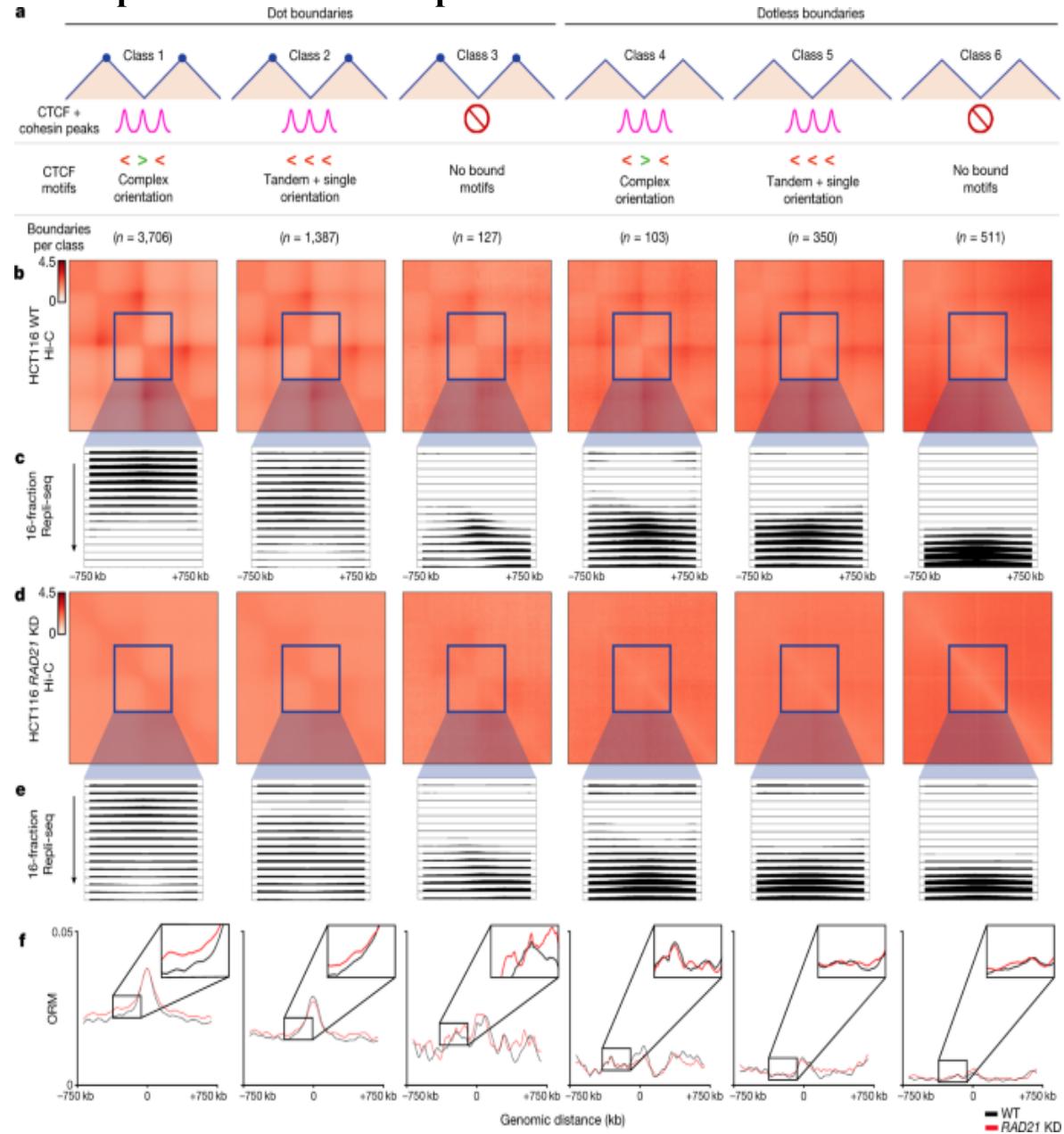
We sought to independently verify our observed link between IZs and boundaries with an orthogonal technique for assaying replication origin activity. Small nascent strand sequencing (SNS-seq) identifies approximately 10 origins per 100 kb of the genome and enriches for high-efficiency origins localized in early replicating regions<sup>[30](#)</sup>. A previous report using ENCODE (Encyclopedia of DNA Elements) phase I pilot microarray data of 1% of the human genome reported enrichment of the cohesin subunit RAD21 at approximately 300 replication origins<sup>[27](#)</sup>. Here, using genome folding features from high-resolution Hi-C data, we find that SNS-seq data from human ES cells<sup>[30](#)</sup> exhibits heightened origin enrichment specifically at class 1 dot boundaries (Extended Data Fig. [5e](#)). Thus, through two independent replication mapping techniques, we observe a strong enrichment of high-efficiency, early-S-phase IZs at a subset of genetically encoded corner-dot TAD/subTAD boundaries. The colocalization of IZs with TAD boundaries generally has been further confirmed recently with super-resolution imaging<sup>[31](#)</sup>.

Transcription correlates with origin placement and efficiency<sup>15,17,32,33,34,35</sup>. To ascertain whether transcription at boundaries could explain our results, we stratified dot boundaries with a complex CTCF orientation (class 1), dotless boundaries with a complex CTCF orientation (class 4) and dotless boundaries with no CTCF occupancy (class 6) into those that also had transcribed genes and those that were devoid of genes or had only inactive genes (Extended Data Fig. 6 and Supplementary Table 8). Boundaries with transcribed genes in the absence of the dot features (Extended Data Fig. 6b) or in the absence of CTCF + cohesin (Extended Data Fig. 6c) did not exhibit precise localization of high-efficiency early IZs. These results are consistent with the literature, as a large proportion of active promoters are not sites of efficient replication initiation, suggesting that further distinguishing features encode human origins<sup>36</sup>. It is also particularly noteworthy that we see enrichment of early IZs at dot boundaries with a complex CTCF motif orientation only when transcribed genes were also present (Extended Data Fig. 6a). Our data suggest that transcription alone is not sufficient to localize high-efficiency early IZs at boundaries. Transcription may cooperate with CTCF and cohesin-based loop extrusion to position high-efficiency IZs replicating in early S phase.

To understand whether cohesin and TAD/subTAD structural integrity are functionally necessary for origin placement in S phase, we examined IZs after global genome folding disruption using wild-type HCT116 cells engineered to degrade the cohesin subunit RAD21 within hours using a degron<sup>23</sup>. Such a system is uniquely suited to test the role of cohesin-mediated extrusion on IZs decoupled from transcription, as only hours of RAD21 degradation results in genome-wide ablation of nearly all loops with minimal short-term effect on transcription<sup>23</sup>. We synchronized HCT116 RAD21–mAID cells in mitosis, degraded RAD21 with auxin throughout G1, and then assessed replication initiation across S phase (Extended Data Fig. 7 and [Supplementary Methods](#)). We identified the same dot and dotless TADs/subTADs and boundary classes in Hi-C from wild-type HCT116 (untreated HCT116 RAD21–mAID) cells as in human ES cells (Fig. 2a and Supplementary Tables 9–14). Consistent with previous reports<sup>23</sup>, our observations show that nearly all dot and dotless boundaries were destroyed following short-term cohesin knockdown in HCT116 cells (Fig. 2b,d and Extended Data Fig. 8). Therefore, although the molecular composition of

boundaries influences their structural features of insulation strength and corner-dot presence, most are dependent on cohesin.

**Fig. 2: Loss of cohesin-mediated TADs/subTADs severely disrupts the genomic placement of DNA replication IZs.**



**a**, Boundary classification in HCT116 wild-type (untreated HCT116 RAD21–mAID) cells conducted as in human ES cells (Fig. 1d) with boundary counts as listed in Supplementary Table 14. The boundary numbers in the figure are provided for autosomal chromosomes alone. **b,d**,

Aggregate peak analysis of the Hi-C observed/expected average interaction frequency of the domains centred on each boundary classification in HCT116 wild-type (WT; untreated HCT116 RAD21–mAID; **b**) and HCT116 *RAD21*-knockdown (KD; auxin-treated HCT116 RAD21–mAID; **d**) cells after cohesin degradation with auxin treatment. The Hi-C source data are from ref. [23](#). **c,e**, High-resolution 16-fraction Repli-seq data in wild-type HCT116 (WT; untreated HCT116 RAD21–mAID; **c**) and HCT116 *RAD21*-knockdown (KD; auxin-treated HCT116 RAD21–mAID; **e**) cells. Each row represents a temporal fraction from S phase, with 16 rows/fractions in total. The Repli-seq signal plotted represents an average across all boundaries in a particular class for that fraction (y-axis) in 50-kb bins across a ±750-kb genomic distance centred on the midpoint of the boundaries (x-axis). Sample sizes for each class are shown in **a**. **f**, ORM data for wild-type (untreated HCT116 RAD21–mAID; black) and *RAD21*-knockdown (auxin-treated HCT116 RAD21–mAID; red) cells.

Previous studies have reported that replication timing domains are not globally altered following genome-wide disruption of cohesin-mediated loops<sup>[37,38,39](#)</sup>. Analyses in these studies relied on the log ratio of DNA synthesized in the first or second halves of S phase (two-fraction early/late Repli-seq)<sup>[40](#)</sup>, the resolution of which renders it difficult to discern IZs. Moreover, previously published two-fraction Repli-seq signals were often quantile normalized<sup>[37,39](#)</sup>, which obscures the localized disruption in IZ placement and timing shifts at specific TAD/subTAD boundaries. We generated and analysed high-resolution 16-fraction Repli-seq data (Fig. [2c,e](#) and Supplementary Table [15](#)), as well as single-molecule optical replication mapping (ORM) data<sup>[17](#)</sup> (Fig. [2f](#)), in both wild-type and cohesin-knockdown HCT116 cells (Extended Data Fig. [7](#) and [Supplementary Methods](#)). As in human ES cells, we observed that 16-fraction Repli-seq data exhibit focal enrichment of high-efficiency/early IZs specifically at dot boundaries marked by CTCF + cohesin co-bound motifs in a complex orientation in wild-type HCT116 cells (class 1; Fig. [2c](#)). Enrichment of early IZs occurs only at boundaries that colocalize with cohesin (Extended Data Fig. [9](#)). Moreover, as in human ES cells, low-efficiency, late IZs were enriched at weak dotless boundaries in wild-type HCT116 cells (Fig. [2c](#)). Using single-molecule ORM data, which can directly assess IZ efficiency as the percentage of molecules that initiate within a particular IZ, we detected

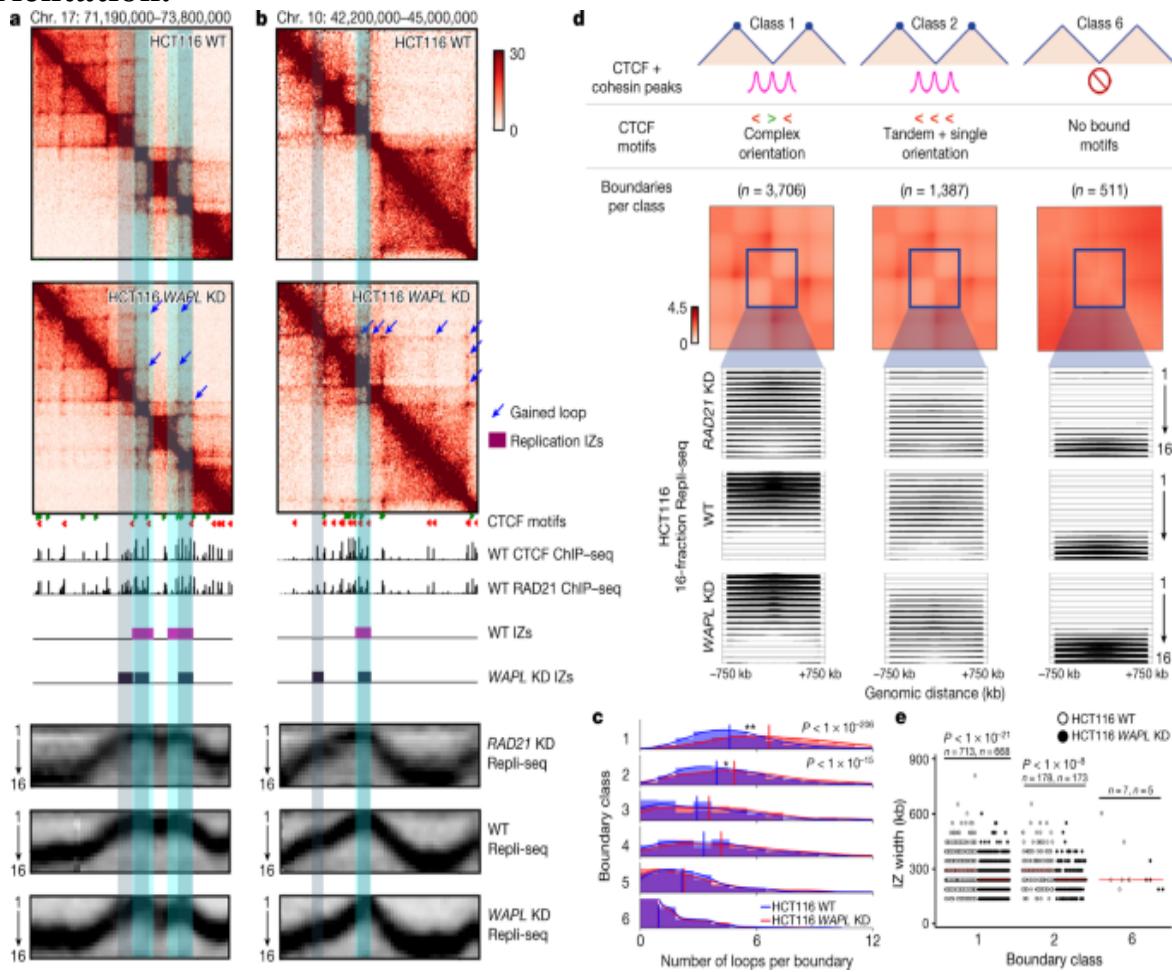
enriched origin initiation specifically at class 1 boundaries (Fig. 2f). Together, our single-molecule and ensemble replication initiation data indicate that early-S-phase IZs fire at a key subset of genetically encoded dot boundaries.

Following ablation of cohesin-mediated boundaries (Fig. 2b,d and Extended Data Fig. 8), we observe severe disruption of high-efficiency early-S-phase IZs specifically at class 1 boundaries, as evidenced by a diffuse and delocalized Repli-seq signal (class 1; Fig. 2c,e). Consistent with our qualitative observations, early wave IZs were less numerous and increased in width specifically at dot boundaries with a complex CTCF motif orientation after loss of cohesin (Extended Data Fig. 10 and Supplementary Table 16). We also noticed that low-efficiency IZs shift to replicating at the end of S phase (fractions 14–16) at dotless boundaries following cohesin knockdown (classes 4–6, Fig. 2c,e and Extended Data Fig. 10). Independently conducted ORM analyses confirmed our observations of IZ disruption by cohesin removal (Fig. 2f). Cell cycle progression and 5-bromodeoxyuridine incorporation was not substantially affected by *RAD21* knockdown<sup>39</sup> (Extended Data Fig. 7). Together, our ensemble and single-molecule IZ data demonstrate that disruption of cohesin-mediated loops during G1 alters the genomic placement where origins or clusters of origins fire during early S phase.

On the basis of our observations, we reason that a failure of cohesin to unload, and therefore the creation of new long-range loops due to more cohesin molecules stalled at complex CTCF boundaries in G1 phase, might result in an increased number of high-efficiency origins or a narrowing of their genomic placement in S phase. Recently, it was reported that knockdown of the gene encoding the cohesin unloading factor WAPL results in increased long-range loops<sup>41</sup>. We examined the genomic placement of IZs in S phase with 16-fraction Repli-seq in wild-type HCT116 cells engineered with an improved degron system (AID2) to degrade WAPL throughout G1 phase<sup>42</sup>. First, we created Hi-C libraries in wild-type and *WAPL*-knockdown HCT116 cells (Fig. 3a,b and Extended Data Fig. 7). Consistent with published results, our observations show that dots indicative of loops are more numerous, and traverse a longer genomic distance, compared with those in wild-type HCT116 cells (Fig. 3a,b and Supplementary Table 18).

We observed that the gain-of-looping phenotype following *WAPL* knockdown occurs most strongly at dot boundaries with a complex CTCF motif orientation (class 1; Fig. 3c). At class 1 boundaries, we observe that early IZs become significantly narrower following *WAPL* knockdown (Fig. 3d,e and Supplementary Table 17). We note that IZs tighten and refine following gain of looping in the *WAPL*-knockdown condition at the same boundaries where IZs grow more diffuse following cohesin knockdown (Fig. 3a,b and Extended Data Fig. 10). Together, the findings from our gain and loss of structural boundary experiments further support a model in which cohesin-based loop extrusion in interphase deterministically informs the placement of the subset of origins that fire during S phase.

**Fig. 3: Gain of looping with *WAPL* degradation narrows the genomic placement of early IZs at dot boundaries with a complex CTCF motif orientation.**

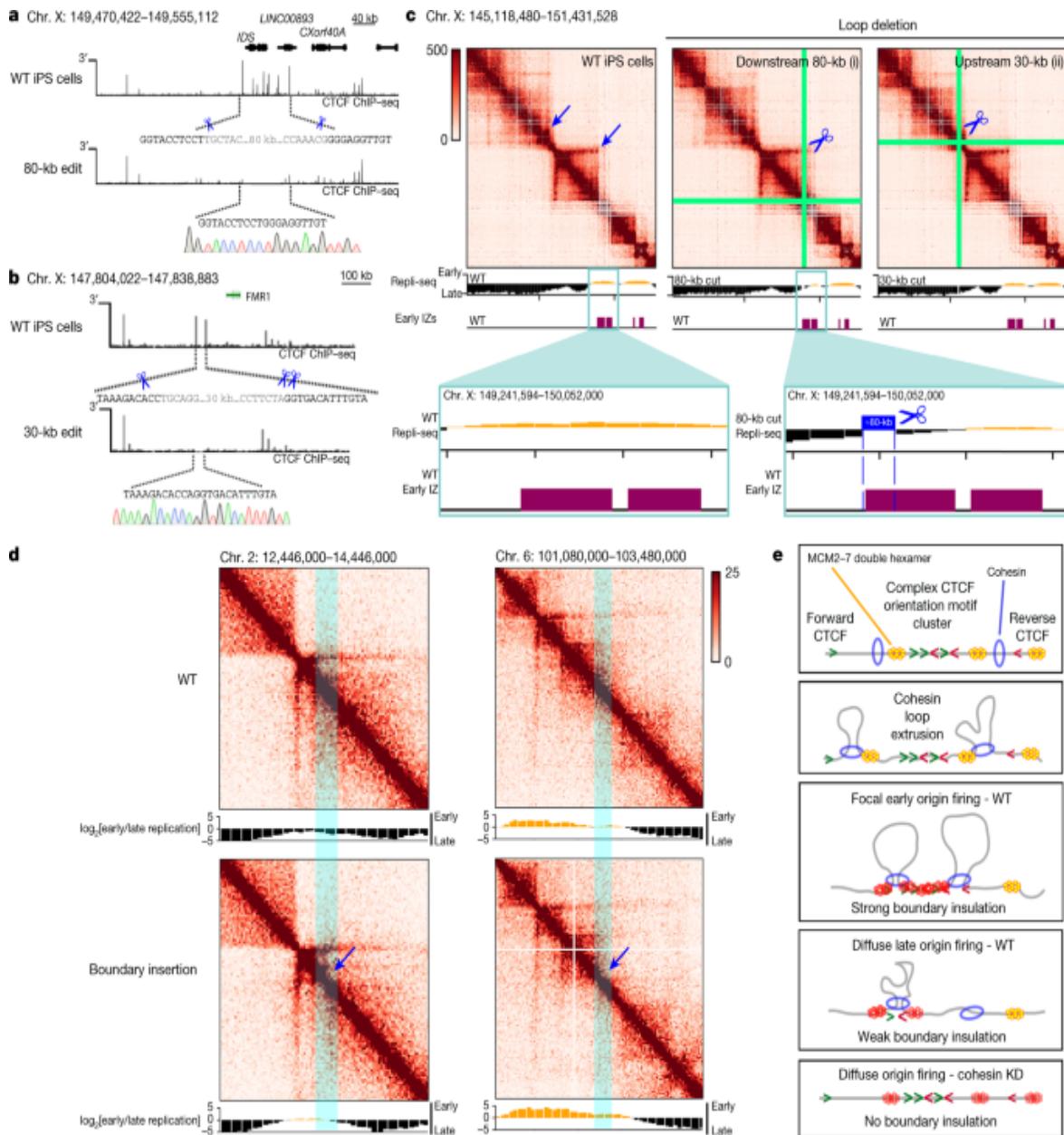


**a,b**, Hi-C maps from wild-type HCT116 (WT; untreated HCT116 WAPL-mAID2) and HCT116 *WAPL*-knockdown (KD; auxin-treated HCT116 WAPL-mAID2) cells for the loci chromosome 17: 71.19–73.8 Mb, hg38 (**a**) and chromosome 10: 42.2–45 Mb, hg38 (**b**). The tracks show CTCF motifs, CTCF ChIP-seq, RAD21 ChIP-seq, high-resolution 16-fraction Repli-seq and IZs. **c**, Distribution of loops per boundary for each of the six boundary classes. Vertical lines demarcate mean number of loops per boundary within each sample and boundary class. Two-tailed Mann–Whitney *U*-test between HCT116 *WAPL*-knockdown and HCT116 wild-type cells for class 1  $P = 2.0 \times 10^{-207}$  and class 2  $P = 1.2 \times 10^{-16}$ . **d**, Averaged Repli-seq for each of the 16 fractions in a  $\pm 750$ -kb window at boundary classes 1, 2 and 6 as detailed in the [Supplementary Methods](#) and Supplementary Table 14. Boundary numbers are provided in the figure for autosomal chromosomes alone. Each Repli-seq row represents a temporal fraction from S phase, there are 16 rows/fractions, and the Repli-seq signal plotted represents an average across all boundaries in a particular class for that fraction (y-axis) in 50-kb bins across a  $\pm 750$ -kb genomic distance centred on the midpoint of boundaries (x-axis). **e**, Width of all IZs colocalized with boundary classes 1, 2 and 6. Two-tailed Mann–Whitney *U* comparing HCT116 wild-type to HCT116 *WAPL*-knockdown cells for class 1  $P = 3.0 \times 10^{-22}$  and class 2  $P = 3.3 \times 10^{-9}$ .

We finally sought to understand whether specific boundaries are necessary and sufficient to regulate IZ firing. We used targeted CRISPR–Cas9 genome editing to delete an 80-kb section of the genome containing a complex array of more than 10 CTCF + cohesin-binding sites with complex motif orientations anchoring a long-range chromatin loop that separates late from early replication timing domains (Fig. 4a). The loop anchor was chosen because it also partially overlaps an early-S-phase IZ, but does not encompass the full IZ, thus allowing us to ablate the loop while keeping much of the IZ intact. We observed a striking local delay of replication timing from early to late following deletion of the 80-kb loop anchor, consistent with the loss of an early IZ (Fig. 4a,c(i)). As a negative control, we deleted a different 30-kb loop anchored by two tandemly oriented CTCF-binding sites within an adjacent late replication timing domain, but not overlapping an IZ (Fig. 4b). Deletion of this 30-kb loop anchor disrupted the dot boundary but preserved the timing and genomic location of DNA

replication (Fig. 4b,c(ii)). The direct overlap of IZs with boundaries precludes our ability to fully decouple them, and overlap of functional elements remains a technical challenge for functional perturbative studies in the genome biology field at large. Nevertheless, our data provide evidence that replication at a specific early IZ can undergo a striking shift to late S phase following ablation of a boundary. These data are consistent with our cohesin-knockdown observations and our model in which boundaries marked by a complex CTCF motif orientation inform the precise placement of high-efficiency IZs.

**Fig. 4: Targeted perturbation leading to gain and loss of structural boundaries can deterministically shift replication timing from early to late S phase.**



**a**, Schematic showing a CRISPR-mediated 80-kb deletion encompassing the *IDS* gene (coordinates of deletion: hg38, chromosome X: 149,470,422–149,555,112). **b**, Schematic showing a CRISPR-mediated 30-kb deletion encompassing two CTCF sites approximately 100 kb upstream from the *FMR1* gene (coordinates of deletion: hg38, chromosome X: 147,804,022–147,838,883). CTCF ChIP-seq tracks for wild-type induced pluripotent stem (iPS) cells and the edited clone are shown. Scissors represent the location of the cut sites verified with Sanger sequencing. **c**, 5C heatmaps (chromosome X: 145,118,480–151,431,528, hg38) and two-fraction Repli-seq tracks in

wild-type iPS cells, and iPS cells with an 80-kb (i) or a 30-kb (ii) loop anchor deletion. Tracks for IZs in human ES cells are overlaid. **d**, Gain of boundary Hi-C and Repli-seq at chromosome 2:  $\approx 13M$  (hg19) and chromosome 6:  $\approx 102M$  (hg19) in HAP1 cells with a transposon inserted boundary<sup>43</sup> and replication timing. **e**, Model of DNA replication initiation determined by high-likelihood cohesin extrusion stalling against strong TAD/subTAD boundaries created high-density arrays of CTCF + cohesin-bound motifs with a complex orientation (early replicating IZs) or low-likelihood cohesin pausing against weak TAD/subTAD boundaries formed by single CTCF motifs (late replicating IZs). Yellow double hexamers depict the MCM2–7 complex at licensed origins. Red double hexamers represent the subset of licensed origins that are activated.

As the direct overlap of IZs with boundaries is not amenable to clean, single-variable ‘loss-of-structure’ perturbative experiments, we also examined a ‘gain-of-structure’ approach in which we assessed whether the introduction of an engineered ectopic boundary was sufficient to induce changes in replication initiation. We mapped replication with two-fraction Repli-seq in published HAP1 cell lines in which we have previously demonstrated a gain in boundary following insertion of an established 2 kb-sized cell-type-invariant boundary element<sup>43</sup>. We observed a striking shift from late to early replication directly at the location of the engineered boundary (Fig. 4d), consistent with the possibility that boundaries can be sufficient for de novo early IZ firing. Together, our data reveal that both global and local gain and loss of structural boundaries can deterministically influence the placement of IZs.

It is well established that the initiation of DNA replication involves two mutually exclusive steps<sup>1,2</sup>. The first step, origin licensing, begins in telophase with the loading of two copies of the mini-chromosome maintenance (MCM2–7) complex<sup>2,44</sup>. MCM2–7 is initially loaded in excess at tens of thousands of sites across the human genome in an inactive form as a double hexamer that encircles double stranded DNA (yellow double hexamers in Fig. 4e). The second step, origin activation, occurs at the onset of S phase. Origin activation involves mechanisms that both prevent further MCM loading and recruit multiple extra factors to initiate the unwinding of the double helix and DNA synthesis<sup>2,44</sup>. In mammalian systems, a critical

mystery remains regarding the mechanisms that governing the selection of a subset of MCM-bound, licensed origins for activation in S phase.

Here we propose a model in which cohesin-mediated loop extrusion and stalling at dot boundaries marked by CTCF + cohesin-binding sites oriented in convergent and divergent directions is required for the positioning of high-efficiency replication origins (Fig. 4e). We propose two possible models to explain the strong localization of high-efficiency IZs to a subset of cohesin-dependent, genetically encoded boundaries: cohesin could directly push licensed MCM double hexamers or other origin activation cofactors along the genome before stalling at high-density arrays of CTCF + cohesin-bound motifs in complex orientations; alternatively, cohesin might pass over many licensed, MCM-bound origins and selectively participate in the activation of those already loaded at boundaries. We also posit that low-efficiency IZs might fire at weaker dotless boundaries later in S phase because cohesin only temporarily pauses during its traversal along the genome, and thus cannot aggregate initiation activity (Fig. 4e). In the cell types from our study, cohesin-mediated loop extrusion is required for IZ placement, and the changes in replication timing are subtle and indirect owing to the altered distance of nearby genomic regions to the nearest initiation site. We note that although we do not see evidence for a dominant role for cohesin on the larger replication timing program, we cannot rule out that cohesin knockdown might have a more profound effect on the replication timing program in other cell types, species and experimental designs.

Previous studies using mass spectrometry and co-immuno-precipitation have reported the direct binding of cohesin to DNA replication factors, such as MCM7, MCM6, MCM4, RFC1 and DNA polymerase  $\alpha$ <sup>27,45</sup>. The MCM complex has the ability to slide after loading and can be pushed by polymerase during transcription<sup>46,47,48</sup>. However, the extent and rate at which this occurs on chromatin in the presence of nucleosomes ( $\approx 11$  nm) is still an open question. The internal diameter of cohesin is 40 nm, whereas the MCM2–7 double hexamer is only 15 nm. The findings of a recent Hi-C and imaging study suggest that, despite their small size, MCM complexes could also serve as boundaries to block cohesin-based loop extrusion<sup>49</sup>. TAD boundaries and loops persist through S phase<sup>50</sup>, but MCMs are removed

from chromatin after IZs fire<sup>1,2</sup>. Therefore, we favour a model in which cohesin pushes licensed MCMs in G1, leading to the localization and activation of a key subset of origins at boundaries with a complex CTCF motif orientation in S phase (Fig. 4e). Nevertheless, both proposed models remain exciting areas for future mechanistic dissection.

Understanding the structure–function relationship of the human genome remains a major challenge for human geneticists and chromatin biologists. Here we stratify TADs and subTADs by their structural and molecular features. We conduct global and local perturbative studies to reveal that genetically encoded TAD/subTAD boundaries formed by cohesin-mediated loop extrusion in G1/pre-S functionally inform genome function in the case of the initiation of DNA replication in S phase. Our work sheds light on the question of whether and how the location of fired origins is deterministically encoded in humans by the genome, epigenome and higher-order chromatin folding.

## Reporting summary

Further information on research design is available in the [Nature Research Reporting Summary](#) linked to this paper.

## Data availability

All new raw data created in this manuscript have been uploaded to the 4D Nucleome portal and will be freely released for full distribution to the public (see specific details below). Processed data files for all figures and extended data figures are provided as Supplementary Tables 1–19. ORM data have been uploaded to the National Center for Biotechnology Information, BioProject database accession number [PRJNA788726](#) (<http://genome.ucsc.edu/s/dsaulebe/ORM%20data%20HCT116>). Two-fraction Repli-seq data for Blobel engineered lines (raw data and processed  $\log_2[\text{early}/\text{late}]$  from three conditions) were obtained from <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE190117>.

Group 1 data (16-fraction Repli-seq data for H1 human ES cells) are available from the 4D Nucleome portal as follows: H1 human ES raw fastq,

<https://data.4dnucleome.org/experiment-sets/4DNESXRBILXJ/>; H1 human ES read-depth-normalized array for visualization,  
<https://data.4dnucleome.org/files-processed/4DNFIEYFQ7C/>; H1 human ES scaled, read-depth-normalized array for IZ calls,  
<https://data.4dnucleome.org/files-processed/4DNFI3N8GHKR/>; H1 human ES early, early–mid and late IZs on read-depth-normalized array,  
<https://data.4dnucleome.org/files-processed/4DNFIRF7WZ3H/>.

Group 2 data (16-fraction Repli-seq data for wild-type HCT116 cells) are available from the 4D Nucleome portal as follows: wild-type HCT116 raw fastq, <https://data.4dnucleome.org/experiment-sets/4DNESNGZM5FG/>; wild-type HCT116 mitochondria-normalized array for IZ calls, <https://data.4dnucleome.org/files-processed/4DNFIPIQTMJ9/>; wild-type HCT116 early, early–mid and late IZs on mitochondria-normalized array, <https://data.4dnucleome.org/files-processed/4DNFI95K53YS/>.

Group 3 data (16-fraction Repli-seq data for wild-type and cohesin-knockdown HCT116 pairing) are available from the 4D Nucleome portal as follows: *RAD21*-knockdown HCT116 raw, <https://data.4dnucleome.org/experiment-sets/4DNES92AU9JR/>; *RAD21*-knockdown HCT116 read-depth-normalized downsampled array for IZ calls, <https://data.4dnucleome.org/files-processed/4DNFI3ZMWG5T/>; *RAD21*-knockdown HCT116 early, early–mid and late IZs called on the read-depth-normalized downsampled array, <https://data.4dnucleome.org/files-processed/4DNFIGOMS9G7/>; wild-type HCT116 raw fastq, <https://data.4dnucleome.org/experiment-sets/4DNESNGZM5FG/>; wild-type HCT116 read-depth-normalized downsampled array for IZ calls, <https://data.4dnucleome.org/files-processed/4DNFI6NGWNOG/>; wild-type HCT116 early, early–mid and late IZs called on the read-depth-normalized downsampled array, <https://data.4dnucleome.org/files-processed/4DNFIYO3H24N/>.

Group 4 data (16-fraction Repli-seq data for wild-type and *WAPL*-knockdown HCT116 pairing) are available from the 4D Nucleome portal as follows: *WAPL*-knockdown HCT116 raw, <https://data.4dnucleome.org/experiment-sets/4DNES72NE7SL/>; *WAPL*-knockdown HCT116 read-depth-normalized downsampled array for IZ calls, <https://data.4dnucleome.org/files-processed/4DNFI7MI88QR/>; *WAPL*-

knockdown HCT116 early, early–mid and late IZs called on the read-depth-normalized downsampled array, <https://data.4dnucleome.org/files-processed/4DNFIDI1QJVA/>; wild-type HCT116 raw fastq, <https://data.4dnucleome.org/experiment-sets/4DNEESNGZM5FG/>; wild-type HCT116 read-depth-normalized downsampled array for IZ calls, <https://data.4dnucleome.org/files-processed/4DNFI6NGWNOG/>; wild-type HCT116 early, early–mid and late IZs called on the read-depth-normalized downsampled array, <https://data.4dnucleome.org/files-processed/4DNFILNNSFMD/>.

Group 5 data (16-fraction Repli-seq data visualization) are available from the 4D Nucleome portal as follows: wild-type HCT116 read-depth-normalized downsampled array for visualization, <https://data.4dnucleome.org/files-processed/4DNFI6NGWNOG/>; *RAD21*-knockdown HCT116 read-depth-normalized downsampled array for visualization, <https://data.4dnucleome.org/files-processed/4DNFI3ZMWG5T/>; *WAPL*-knockdown HCT116 read-depth-normalized downsampled array for visualization, <https://data.4dnucleome.org/files-processed/4DNFI7MI88QR/>.

Hi-C data for wild-type and *WAPL*-knockdown HCT116 pairing are available from the 4D Nucleome portal as follows: *WAPL*-knockdown HCT116 raw Hi-C, <https://data.4dnucleome.org/experiment-set-replicates/4DNEES1JP4KZ1/>; *WAPL*-knockdown HCT116 normalized balanced Hi-C matrices, <https://data.4dnucleome.org/files-processed/4DNFIY5939F3/>; *WAPL*-knockdown HCT116 loops, <https://data.4dnucleome.org/files-processed/4DNFILP7BD5H/>; wild-type HCT116 raw Hi-C, <https://data.4dnucleome.org/experiment-set-replicates/4DNESNSTBMBY/>; wild-type HCT116 normalized balanced Hi-C matrices, <https://data.4dnucleome.org/files-processed/4DNFI5MR78O6/>; wild-type HCT116 loops, <https://data.4dnucleome.org/files-processed/4DNFIOQLL854/>.

Two-fraction Repli-seq data for human iPS wild-type and two CRISPR-engineered lines (raw data and processed log<sub>2</sub>[early/late] from three conditions) are available from the 4D Nucleome portal as follows: wild-type human iPS line raw data, <https://data.4dnucleome.org/experiment-sets/4DNEESNGZM5FG/>.

[sets/4DNEYES9QD/](https://data.4dnucleome.org/experiment-sets/4DNEYES9QD/); wild-type human iPS line log<sub>2</sub>[early/late],  
<https://data.4dnucleome.org/files-processed/4DNFI5WEY784/>; human engineered clone 1 80-kb-IZ-deletion iPS line raw data,  
<https://data.4dnucleome.org/experiment-sets/4DNESE3WCUAQ/>; human engineered clone 1 80-kb-IZ-deletion iPS line log<sub>2</sub>[early/late],  
<https://data.4dnucleome.org/files-processed/4DNFIZMB415V/>; human engineered clone 2 30-kb-control-deletion iPS line raw data,  
<https://data.4dnucleome.org/experiment-sets/4DNES66YWJU7/>; human engineered clone 2 30-kb-control-deletion iPS line log<sub>2</sub>[early/late],  
<https://data.4dnucleome.org/files-processed/4DNFIWDMF7HW/>.

5C data for human IPS wild-type and two engineered lines (primer bed file, raw heatmaps and processed heatmaps from three conditions) are available from the 4D Nucleome portal as follows: wild-type human iPS line raw data, <https://data.4dnucleome.org/experiment-set-replicates/4DNESLRDUPZ6/>; wild-type human iPS line balanced 5C data, replicate 1, <https://data.4dnucleome.org/files-processed/4DNFIXM8V3ZB/>, replicate 2, <https://data.4dnucleome.org/files-processed/4DNFIDB6M1ZN/>; wild-type human engineered clone 1 80-kb-boundary-deletion iPS line raw data, <https://data.4dnucleome.org/experiment-set-replicates/4DNES39F1QWU/>; wild-type human engineered clone 1 80-kb-boundary-deletion iPS line balanced 5C data, <https://data.4dnucleome.org/files-processed/4DNFIA8P94BX/>; wild-type human engineered clone 2 30-kb-control-deletion iPS line raw data, <https://data.4dnucleome.org/experiment-set-replicates/4DNES3PDMUHG/>; wild-type human engineered clone 2 30-kb-control-deletion iPS line balanced 5C data: replicate 1, <https://data.4dnucleome.org/files-processed/4DNFI7WZYRHP/>, replicate 2, <https://data.4dnucleome.org/files-processed/4DNFI7V4VXAQ/>.

## Code availability

We freely release all custom code for loop, TAD and subTAD detection at the following bitbucket links: TAD/subTAD detection, [https://bitbucket.org/creminslab/cremins\\_lab\\_tadsubtad\\_calling\\_pipeline\\_11\\_6\\_2021/](https://bitbucket.org/creminslab/cremins_lab_tadsubtad_calling_pipeline_11_6_2021/); loop detection, [https://bitbucket.org/creminslab/cremins\\_lab\\_loop\\_calling\\_pipeline\\_11\\_6\\_2021/src/initial/](https://bitbucket.org/creminslab/cremins_lab_loop_calling_pipeline_11_6_2021/src/initial/).

## References

1. Bellush, J. M. & Whitehouse, I. DNA replication through a chromatin environment. *Philos. Trans. R. Soc. B* **372**, 20160287 (2017).
2. Mechali, M. Eukaryotic DNA replication origins: many choices for appropriate answers. *Nat. Rev. Mol. Cell Biol.* **11**, 728–738 (2010).
3. Dixon, J. R. et al. Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485**, 376–380 (2012).
4. Nora, E. P. et al. Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature* **485**, 381–385 (2012).
5. Hou, C., Li, L., Qin, Z. S. & Corces, V. G. Gene density, transcription, and insulators contribute to the partition of the *Drosophila* genome into physical domains. *Mol. Cell* **48**, 471–484 (2012).
6. Sexton, T. et al. Three-dimensional folding and functional organization principles of the *Drosophila* genome. *Cell* **148**, 458–472 (2012).
7. Phillips-Cremins, J. E. et al. Architectural protein subclasses shape 3D organization of genomes during lineage commitment. *Cell* **153**, 1281–1295 (2013).
8. 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).
9. Krietenstein, N. et al. Ultrastructural details of mammalian chromosome architecture. *Mol. Cell* **78**, 554–565 (2020).
10. Hsieh, T. S. et al. Resolving the 3D landscape of transcription-linked mammalian chromatin folding. *Mol. Cell* **78**, 539–553 (2020).
11. Akgol Oksuz, B. et al. Systematic evaluation of chromosome conformation capture assays. *Nat. Methods* **18**, 1046–1055 (2021).

12. Fudenberg, G. et al. Formation of chromosomal domains by loop extrusion. *Cell Rep.* **15**, 2038–2049 (2016).
13. Vian, L. et al. The energetics and physiological impact of cohesin extrusion. *Cell* **173**, 1165–1178 (2018).
14. Sanborn, A. L. et al. Chromatin extrusion explains key features of loop and domain formation in wild-type and engineered genomes. *Proc. Natl Acad. Sci. USA* **112**, E6456 (2015).
15. Petryk, N. et al. Replication landscape of the human genome. *Nat. Commun.* **7**, 10208 (2016).
16. Zhao, P. A., Sasaki, T. & Gilbert, D. M. High-resolution Repli-Seq defines the temporal choreography of initiation, elongation and termination of replication in mammalian cells. *Genome Biol.* **21**, 76 (2020).
17. Wang, W. et al. Genome-wide mapping of human DNA replication by optical replication mapping supports a stochastic model of eukaryotic replication. *Mol. Cell* **81**, 2975–2988 (2021).
18. Ryba, T. et al. Evolutionarily conserved replication timing profiles predict long-range chromatin interactions and distinguish closely related cell types. *Genome Res.* **20**, 761–770 (2010).
19. Pope, B. D. et al. Topologically associating domains are stable units of replication-timing regulation. *Nature* **515**, 402–405 (2014).
20. Norton, H. K. et al. Detecting hierarchical genome folding with network modularity. *Nat. Methods* **15**, 119–122 (2018).
21. Gilgenast, T. G. & Phillips-Cremins, J. E. Systematic evaluation of statistical methods for identifying looping interactions in 5C data. *Cell Syst.* **8**, 197–211 (2019).
22. Fernandez, L. R., Gilgenast, T. G. & Phillips-Cremins, J. E. 3DeFDR: statistical methods for identifying cell type-specific looping interactions in 5C and Hi-C data. *Genome Biol.* **21**, 219 (2020).

23. Rao, S. S. P. et al. Cohesin loss eliminates all loop domains. *Cell* **171**, 305–320 (2017).
24. Schwarzer, W. et al. Two independent modes of chromatin organization revealed by cohesin removal. *Nature* **551**, 51–56 (2017).
25. Davidson, I. F. et al. Rapid movement and transcriptional re-localization of human cohesin on DNA. *EMBO J.* **35**, 2671–2685 (2016).
26. Pherson, M., Misulovin, Z., Gause, M. & Dorsett, D. Cohesin occupancy and composition at enhancers and promoters are linked to DNA replication origin proximity in *Drosophila*. *Genome Res.* **29**, 602–612 (2019).
27. Guillou, E. et al. Cohesin organizes chromatin loops at DNA replication factories. *Genes Dev.* **24**, 2812–2822 (2010).
28. de Wit, E. et al. CTCF binding polarity determines chromatin looping. *Mol. Cell* **60**, 676–684 (2015).
29. Tang, Z. et al. CTCF-mediated human 3D genome architecture reveals chromatin topology for transcription. *Cell* **163**, 1611–1627 (2015).
30. Besnard, E. et al. Unraveling cell type-specific and reprogrammable human replication origin signatures associated with G-quadruplex consensus motifs. *Nat. Struct. Mol. Biol.* **19**, 837–844 (2012).
31. Li, Y. et al. Transcription-coupled structural dynamics of topologically associating domains regulate replication origin efficiency. *Genome Biol.* **22**, 206 (2021).
32. Sequeira-Mendes, J. et al. Transcription initiation activity sets replication origin efficiency in mammalian cells. *PLoS Genet.* **5**, e1000446 (2009).
33. Chen, Y. H. et al. Transcription shapes DNA replication initiation and termination in human cells. *Nat. Struct. Mol. Biol.* **26**, 67–77 (2019).

34. Cayrou, C. et al. The chromatin environment shapes DNA replication origin organization and defines origin classes. *Genome Res.* **25**, 1873–1885 (2015).
35. Liu, Y. et al. Transcription shapes DNA replication initiation to preserve genome integrity. *Genome Biol.* **22**, 176 (2021).
36. Cadoret, J. C. et al. Genome-wide studies highlight indirect links between human replication origins and gene regulation. *Proc. Natl Acad. Sci. USA* **105**, 15837–15842 (2008).
37. Oldach, P. & Nieduszynski, C. A. Cohesin-mediated genome architecture does not define DNA replication timing domains. *Genes* **10**, 196 (2019).
38. Cremer, M. et al. Cohesin depleted cells rebuild functional nuclear compartments after endomitosis. *Nat. Commun.* **11**, 6146 (2020).
39. Sima, J. et al. Identifying cis elements for spatiotemporal control of mammalian DNA replication. *Cell* **176**, 816–830(2019).
40. Hiratani, I. et al. Global reorganization of replication domains during embryonic stem cell differentiation. *PLoS Biol.* **6**, e245 (2008).
41. Haarhuis, J. H. I. et al. The cohesin release factor WAPL restricts chromatin loop extension. *Cell* **169**, 693–707 (2017).
42. Yesbolatova, A. et al. The auxin-inducible degron 2 technology provides sharp degradation control in yeast, mammalian cells, and mice. *Nat. Commun.* **11**, 5701 (2020).
43. Zhang, D. et al. Alteration of genome folding via contact domain boundary insertion. *Nat. Genet.* **52**, 1076–1087 (2020).
44. Dimitrova, D. S., Prokhorova, T. A., Blow, J. J., Todorov, I. T. & Gilbert, D. M. Mammalian nuclei become licensed for DNA replication during late telophase. *J. Cell Sci.* **115**, 51–59 (2002).

45. Ryu, M. J. et al. Direct interaction between cohesin complex and DNA replication machinery. *Biochem. Biophys. Res. Commun.* **341**, 770–775 (2006).
46. Gros, J. et al. Post-licensing specification of eukaryotic replication origins by facilitated Mcm2-7 sliding along DNA. *Mol. Cell* **60**, 797–807 (2015).
47. Powell, S. K. et al. Dynamic loading and redistribution of the Mcm2-7 helicase complex through the cell cycle. *EMBO J.* **34**, 531–543 (2015).
48. Sasaki, T. et al. The Chinese hamster dihydrofolate reductase replication origin decision point follows activation of transcription and suppresses initiation of replication within transcription units. *Mol. Cell. Biol.* **26**, 1051–1062 (2006).
49. Dequeker, B. J. H. et al. MCM complexes are barriers that restrict cohesin-mediated loop extrusion. *Nature* <https://doi.org/10.1038/s41586-022-04730-0> (2022).
50. Nagano, T. et al. Cell-cycle dynamics of chromosomal organization at single-cell resolution. *Nature* **547**, 61–67 (2017).

## Acknowledgements

We thank members of the 4D Nucleome community and the laboratories of J.E.P.-C. and D.M.G. for helpful discussions. J.E.P.-C. is a New York Stem Cell Foundation – Robertson Investigator and an Alfred P. Sloan Foundation Fellow. J.D. is an investigator of the Howard Hughes Medical Institute. This research was supported by National Institute of Mental Health grants (1R01MH120269 and 1DP1MH129957; J.E.P.-C.), 4D Nucleome Common Fund grants (1U01HL12999801, 1U01DK127405 and 1U01DA052715; J.E.P.-C.), a National Science Foundation (NSF) CAREER Award (CBE-1943945; J.E.P.-C.), an NSF Emerging Frontiers in Research Innovation grant (1933400; J.E.P.-C.), 4D Nucleome Common Fund grants (DK107980 and HG011536; J.D.) and National Institutes of Health grants (R01HG010658 and U54DK107965; D.M.G.). The PhD fellowship to D.S.

was provided by the CNRS 80|Prime interdisciplinary programme and W.W. was supported by a COFUND IC-3i International PhD fellowship.

## Author information

### Author notes

1. These authors contributed equally: Daniel J. Emerson, Peiyao A. Zhao, Ashley L. Cook

## Authors and Affiliations

1. Department of Bioengineering, University of Pennsylvania, Philadelphia, PA, USA

Daniel J. Emerson, Ashley L. Cook, R. Jordan Barnett, Chunmin Ge, Linda Zhou, Zoltan Simandi, Miriam K. Minsk, Katelyn R. Titus, Wanfeng Gong & Jennifer E. Phillips-Cremins

2. Epigenetics Institute, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA

Daniel J. Emerson, Ashley L. Cook, R. Jordan Barnett, Chunmin Ge, Linda Zhou, Zoltan Simandi, Miriam K. Minsk, Katelyn R. Titus, Wanfeng Gong & Jennifer E. Phillips-Cremins

3. Department of Genetics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA

Daniel J. Emerson, Ashley L. Cook, R. Jordan Barnett, Chunmin Ge, Linda Zhou, Zoltan Simandi, Miriam K. Minsk, Katelyn R. Titus, Wanfeng Gong & Jennifer E. Phillips-Cremins

4. Department of Biological Science, Florida State University, Tallahassee, FL, USA

Peiyao A. Zhao, Kyle N. Klein & David M. Gilbert

5. Institut Curie, PSL Research University, CNRS UMR3244, Dynamics of Genetic Information, Sorbonne Université, Paris, France

Dalila Saulebekova, Weitao Wang & Chun-Long Chen

6. Children's Hospital of Pennsylvania, Philadelphia, PA, USA

Di Zhang

7. University of Massachusetts Chan Medical School, Worcester, MA, USA

Liyan Yang, Sergey V. Venev, Johan H. Gibcus & Job Dekker

8. Department of Biochemistry and Molecular Genetics, Feinberg School of Medicine, Northwestern University, Chicago, Illinois, USA

Hongbo Yang & Feng Yue

9. San Diego Biomedical Research Institute, San Diego, CA, USA

Takayo Sasaki & David M. Gilbert

10. Department of Chromosome Science, National Institute of Genetics, Research Organization of Information and Systems (ROIS), Mishima, Japan

Masato T. Kanemaki

11. Department of Genetics, The Graduate University for Advanced Studies (Sokendai), Mishima, Japan

Masato T. Kanemaki

12. Howard Hughes Medical Institute, Chevy Chase, MD, USA

Job Dekker

13. New York Stem Cell Foundation Robertson Investigator, New York, NY, USA

Jennifer E. Phillips-Cremins

## Contributions

Conceptualization: D.J.E., P.A.Z., D.M.G., J.E.P.-C. Experimentation: C.G., L.Z., Z.S., K.K., W.G., L.Y., J.H.G., T.S., H.Y., F.Y., J.D., D.M.G., J.E.P.-C. Computation and visualization: D.J.E., P.A.Z., A.L.C., M.K.M., R.J.B., K.R.T., S.V.V., D.S., W.W., C.-L.C., D.M.G., J.E.P.-C. Funding acquisition: J.D., D.M.G., J.E.P.-C. Project administration: D.M.G., J.E.P.-C. Writing: D.J.E., P.A.Z., A.L.C., J.E.P.-C. Review and editing: D.J.E., P.A.Z., A.L.C., M.K.M., J.D., C.-L.C., D.M.G., J.E.P.-C. Critical reagents: D.Z., M.T.K.

## Corresponding author

Correspondence to [Jennifer E. Phillips-Cremins](#).

## Ethics declarations

## Competing interests

The authors declare no competing interests.

## Peer review

## Peer review information

*Nature* thanks Benjamin Rowland and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. [Peer reviewer reports](#) are available.

## Additional information

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

## Extended data figures and tables

### Extended Data Fig. 1 Hi-C and 16-fraction Repli-seq in H1 human ES cells.

Hi-C maps showing **(a)** chr10: 88.5 - 91.0 Mb locus and **(b)** chr10: 108 - 111 Mb locus. Blue lines, Layer 3 most nested TADs/subTADs. Green lines, Layer 2 intermediate TAD/subTAD nesting. Magenta lines, Layer 1 highest-layer non-nested TADs/subTADs. Green rectangles, corner-dots. Tracks show CTCF motifs at colocalized CTCF+cohesin peaks (green, forward; red, reverse), A/B compartments (green, A compartment; red, B compartment), CTCF CUT&RUN and cohesin ChIP-seq (black), low-resolution two-fraction replication timing domains (yellow, early replication timing; black, late replication timing), 16-fraction Repli-seq data, and initiation zones (magenta, all IZs; purple, Early/Early-mid/Late IZs). Genome build, hg38.

### Extended Data Fig. 2 Hi-C and 16-fraction Repli-seq in H1 human ES cells.

Hi-C maps showing (a) chr2: 27 - 30 Mb locus and (b) chr10: 54 - 57 Mb locus. Blue lines, Layer 3 most nested TADs/subTADs. Green lines, Layer 2 intermediate TAD/subTAD nesting. Magenta lines, Layer 1 highest-layer non-nested TADs/subTADs. Green rectangles, corner-dots. Tracks show CTCF motifs at colocalized CTCF+cohesin peaks (green, forward; red, reverse), A/B compartments (green, A compartment; red, B compartment), CTCF CUT&RUN and cohesin ChIP-seq (black), low-resolution two-fraction replication timing domains (yellow, early replication timing; black, late replication timing), 16-fraction Repli-seq data, and initiation zones (magenta, all IZs; purple, Early/Early-mid/Late IZs). Genome build, hg38.

### Extended Data Fig. 3 Hi-C and 16-fraction Repli-seq in H1 human ES cells.

Hi-C maps showing **(a)** chr7: 136 – 138 Mb locus and **(b)** chr10: 84 – 86 Mb locus. Blue lines, Layer 3 most nested TADs/subTADs. Green lines, Layer 2

intermediate TAD/subTAD nesting. Magenta lines, Layer 1 highest-layer non-nested TADs/subTADs. Green rectangles, corner-dots. Tracks show CTCF motifs at colocalized CTCF+cohesin peaks (green, forward; red, reverse), A/B compartments (green, A compartment; red, B compartment), CTCF CUT&RUN and cohesin ChIP-seq (black), low-resolution two-fraction replication timing domains (yellow, early replication timing; black, late replication timing), 16-fraction Repli-seq data, and initiation zones (magenta, all IZs; purple, Early/Early-mid/Late IZs). Genome build, hg38.

### **Extended Data Fig. 4 Loop calling in H1 human ES Hi-C across a series of single-variable parameter changes.**

**(a)** Example genomic locus in human ES cells (chr4: 52.9 - 54.9 Mb, hg38, scale 0–200) with 10 methodological variants of corner-dot detection (Options A through J detailed in the [Supplementary Methods](#)). In teal, we highlight Options D, F, and H as our recommended loop calling parameters in Hi-C 2.5 generated from human ES cells for conservative, intermediate, and permissive calls respectively. Option D – our conservative loop calling set – is indicated by a teal box and was used to call loops for the analysis in the main paper. **(b)** Bar graph showing the number of loops called across autosomes for loop calling parameters Options A through J. **(c)** We computed right-tailed, one-tailed empirical p-values using a randomization test with Early, Early-mid, and Late S phase IZs and size- and A/B compartment-matched null IZs ([Supplemental Methods](#)) across boundaries derived from Options A-J dot calling variants. **(d)** Number of TAD/subTAD boundaries from autosomal chromosomes classified into the following categories: double-dot complex CTCF motif orientation (DD complex), double-dot tandem + single CTCF motif orientation (DD tandem), double-dot no CTCF (DD 0 CTCF), single-dot CTCF complex motif orientation (SD complex), single-dot CTCF tandem + single motif orientation (SD tandem), single-dot no CTCF (SD 0 CTCF), dotless complex CTCF motif orientation (ND complex), dotless tandem + single CTCF motif orientation (ND tandem), and dotless no CTCF (ND 0 CTCF).

### **Extended Data Fig. 5 Statistical test and SNS-seq in H1 human ES cells reveals the enrichment of Early IZs at class 1 boundaries.**

**(a)** Boundary classification schematic in human ES cells with the following boundary counts: (i) N = 4,404, (ii) N = 2,258, (iii) N = 126, (iv) N = 51, (v) N = 320, (vi) N = 346. Boundary class numbers in figure and caption provided for autosomal chromosomes only. **(b-d)** Statistical test computing proximity of IZs to TAD/subTAD boundaries compared to expectation in hES Hi-C autosomes. We computed right-tailed, one-tailed empirical p-values using a randomization test with **(b)** early, **(c)** early-mid, and **(d)** late S phase IZs and size- and A/B compartment-matched null IZs ([Supplemental Methods](#)). Test statistic for real IZs (red line) represents the difference between the average null IZ distance to closest boundary and average real IZ distance to closest boundary (detailed in [Supplemental Methods](#)). Null distribution represents the difference between the average distance to the closest boundary of two reshuffled sets of null IZs. **(e)** We plotted the average SNS-seq signal (reads per million) 500 kb up- and down-stream of the 6 boundary classes. SNS-seq data in human ES cells was acquired from Besnard et al<sup>51</sup>. The overall level of SNS-seq signal at Dot boundaries was also higher than Dotless boundaries, reinforcing the shared propensity of SNS-seq origins and corner-dot TADs/subTADs to both be enriched in the same genomic compartment (A compartment), which we controlled for in our statistical tests.

### [Extended Data Fig. 6 Patterns of 16-fraction Repli-seq in boundaries +/- transcribed genes in H1 human ES cells.](#)

Repli-seq was averaged for each of the 16 fractions in a  $\pm$  750 kb window at **(a)** Boundary class 1, dot boundaries with complex CTCF motif orientation, **(b)** Boundary class 4, dotless boundaries with complex CTCF motif orientation, and **(c)** Boundary class 6, dotless boundaries with no CTCF, further stratified by colocalization with transcribed genes (+ transcription) vs. no genes & no transcribed genes (- transcription) within  $\pm$  100 kb of the midpoint of the boundary. 16 Fraction Repli-seq images pileup scale 0.6–1.85. Boundary numbers provided in figure for autosomal chromosomes only.

### [Extended Data Fig. 7 HCT116 characterization leading to the generation of wild type, WAPL knock-down, and RAD21 knock-down genomics libraries .](#)

**(a)** Treatment and sample collection timeline of HCT116 RAD21-mAID and HCT116 WAPL-mAID2 cells for high-resolution 16-fraction Repli-seq. **(b-c)** Propidium Iodide FACS histograms measuring DNA content for **(b)** HCT116 WAPL-mAID2 cells in asynchronous cultures and immediately after mitotic shake-off conditions, **(c)** auxin-treated HCT116 RAD21-mAID cells and HCT116 WAPL-mAID2 cells at specified time points after mitotic shake-off. No clear defect in cell cycle progression was observed. **(d)** Western blot of RAD21 protein in HCT116 RAD21-mAID cells for untreated control and timepoints after auxin treatment post mitotic shake off. Ponceau S stain for total protein. Blot run on one set of samples. **(e)** Western blot of WAPL protein in HCT116 WAPL-mAID2 cells for untreated control and timepoints after auxin treatment post mitotic shake off. Ponceau S stain for total protein. Blot run on one set of samples. **(f)** Total IP efficiency (genomic DNA over mitochondrial DNA) for each of 16 S phase fractions of high-resolution 16-fraction Repli-seq for HCT116 WT, HCT116 RAD21-mAID KD, and HCT116 WAPL-mAID2 KD cells.

**Extended Data Fig. 8 Insulation score changes at Boundary classes 1–6 in HCT116 cells upon cohesin and WAPL knock-down.**

Average insulation score for each of the six boundary classes in a  $\pm$  760 kb window for wild type HCT116 (WT; untreated HCT116 WAPL-mAID2), HCT116 with cohesin knock-down (RAD21 KD; auxin-treated HCT116 RAD21-mAID), and HCT116 with WAPL knock-down (WAPL KD; auxin-treated HCT116 WAPL-mAID2). The six boundary classes have the following boundary counts (i) N = 3,706 (ii) N = 1,387, (iii) N = 127, (iv) N = 103, (v) N = 350, (vi) N = 511. Boundary numbers provided in figure and caption for autosomal chromosomes only.

**Extended Data Fig. 9 Patterns of 16-fraction Repli-seq in Dot versus Dotless boundaries +/- cohesin in HCT116.**

Repli-seq was averaged for each of the 16 fractions in a  $\pm$  750 kb window at **(a)** all boundaries with demarcating dot TAD/subTADs on one or both sides, Boundary classes 1–3, or **(b)** all boundaries with dotless TADs/subTADs on both sides, Boundary classes 4–6, further stratified by

the colocalization with cohesin ChIP-seq peaks (+ cohesin) vs. no cohesin peaks (- cohesin) within  $\pm$  100 kb of the boundary. 16 Fraction Repli-seq images with cohesin pileup scale 5.0–9.8, no cohesin pileup scale 5.0–13.0. Boundary numbers provided in figure for autosomal chromosomes only.

### **Extended Data Fig. 10 Number and width of IZs in HCT116 cells upon cohesin and WAPL knock-down.**

**(a)** Boundary classes in HCT116 with the following boundary counts (i) N = 3,706 (ii) N = 1,387, (iii) N = 127, (iv) N = 103, (v) N = 350, (vi) N = 511. Boundary numbers provided in figure and caption for autosomal chromosomes only. **(b)** Number of IZs in WT only (unique to wild type HCT116 – white bar), invariant (wild type HCT116 overlapping auxin-treated HCT116 RAD21-mAID KD – middle black bar), and RAD21 KD only (unique to auxin-treated HCT116 RAD21-mAID KD – right black bar). **(c)** Width of all IZs colocalized with Boundary classes 1 - 6 in HCT116 WT and HCT116 RAD21 KD conditions. Median value indicated by red line. Two-tailed Mann-Whitney U comparing overlapping IZs in HCT116 WT and HCT116 RAD21 KD samples. Only IZs overlapping in both HCT116 wild type and HCT116 RAD21 knock-down are plotted. **(d)** Number of IZs in WT only (unique to wild type HCT116 – white bar), invariant (wild type HCT116 overlapping auxin-treated HCT116 WAPL-mAID2 KD – middle black bar), and WAPL KD only (unique to auxin-treated HCT116 WAPL-mAID2 KD – right black bar). **(e)** Width of all IZs colocalized with Boundary classes 1 - 6 in HCT116 WT and HCT116 WAPL KD conditions. Median value indicated by red line. Two-tailed Mann-Whitney U comparing overlapping IZs in HCT116 WT and HCT116 WAPL KD samples. Only IZs overlapping in both HCT116 wild type and HCT116 WAPL knock-down are plotted.

## **Supplementary information**

### **Supplementary Information**

This file contains full legends for Supplementary Tables 1–19, Methods, Fig. 1 and References.

## **Reporting Summary**

## **Supplementary Tables**

This zipped file contains Supplementary Tables 1–19 (see Supplementary Information for full descriptions).

## **Peer Review File**

## **Rights and permissions**

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

## **Reprints and Permissions**

## **About this article**

## **Cite this article**

Emerson, D.J., Zhao, P.A., Cook, A.L. *et al.* Cohesin-mediated loop anchors confine the locations of human replication origins. *Nature* **606**, 812–819 (2022). <https://doi.org/10.1038/s41586-022-04803-0>

- Received: 31 December 2020

- Accepted: 26 April 2022
- Published: 08 June 2022
- Issue Date: 23 June 2022
- DOI: <https://doi.org/10.1038/s41586-022-04803-0>

## Share this article

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

[Get shareable link](#)

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

[Copy to clipboard](#)

Provided by the Springer Nature SharedIt content-sharing initiative

---

This article was downloaded by **calibre** from <https://www.nature.com/articles/s41586-022-04803-0>

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

# Amendments & Corrections

- [\*\*Author Correction: Defining the consequences of genetic variation on a proteome-wide scale\*\*](#) [ 07 Jun 2022]
  -
- [\*\*Publisher Correction: Virtual communication curbs creative idea generation\*\*](#) [ 07 Jun 2022]
  -
- [\*\*Publisher Correction: Genomic reconstruction of the SARS CoV-2 epidemic in England\*\*](#) [ 07 Jun 2022]
  -

- Author Correction
- [Published: 07 June 2022](#)

# Author Correction: Defining the consequences of genetic variation on a proteome-wide scale

- [Joel M. Chick](#),
- [Steven C. Munger](#),
- [Petr Simecek](#),
- [Edward L. Huttlin](#),
- [Kwangbom Choi](#),
- [Daniel M. Gatti](#),
- [Narayanan Raghupathy](#),
- [Karen L. Svenson](#),
- [Gary A. Churchill](#) &
- [Steven P. Gygi](#)

[Nature](#) volume **606**, page E16 (2022)

- 419 Accesses
- 10 Altmetric
- [Metrics details](#)

The [Original Article](#) was published on 15 June 2016

Correction to: *Nature* <https://doi.org/10.1038/nature18270> Published online 15 June 2016

In this Article, the sample identifiers for two consecutive tandem mass tag (TMT) 10-plexes in the proteomics dataset were inadvertently swapped. We corrected the 18 affected sample identifiers for which we could verify their identity by comparison with the RNA-sequencing data and removed two ambiguous samples. We then re-analysed the corrected dataset and compared the protein quantitative trait loci (pQTL) and mediation results with those originally reported. We verified that every specific pQTL and mediator–target interaction highlighted in the original Article replicated successfully in the new analysis with the corrected proteomics dataset. We identify more significant local and distant pQTLs with the corrected dataset. Nearly 97% of reported local pQTLs replicate after correcting for the sample swaps, and we detect 439 additional significant local pQTLs. Distant pQTLs replicated at a much lower rate, consistent with the replication rate we estimated for distant eQTL (Extended Data Fig. 4) and those reported for other eQTL datasets of similar size<sup>1</sup>. Highly significant distant pQTL replicated at much higher rates, and every distant pQTL individually highlighted in the Article successfully replicated in the corrected dataset. Finally, correcting for the sample swaps resulted in more mediator proteins being identified for more distant pQTL. In summary, the results from our re-analysis of the corrected dataset reinforce the conclusions made in the original Article. We have updated Supplementary Tables 3, 5 and 8 (proteomics dataset, list of pQTLs and mediation results, respectively) of the original Article to reflect these corrections (see Supplementary Information to this Amendment). We regret the oversight and are grateful to B. Zhang and colleagues for bringing it to our attention. The original Article has not been corrected online.

## References

1. Strunz, T. et al. A mega-analysis of expression quantitative trait loci (eQTL) provides insight into the regulatory architecture of gene expression variation in liver. *Sci. Rep.* **8**, 5865 (2018).

## Author information

Author notes

1. These authors contributed equally: Joel M. Chick, Steven C. Munger
2. These authors jointly supervised this work: Gary A. Churchill & Steven P. Gygi

## Authors and Affiliations

1. Harvard Medical School, Boston, Massachusetts, USA  
Joel M. Chick, Edward L. Huttlin & Steven P. Gygi
2. The Jackson Laboratory, Bar Harbor, Maine, USA  
Steven C. Munger, Petr Simecek, Kwangbom Choi, Daniel M. Gatti, Narayanan Raghupathy, Karen L. Svenson & Gary A. Churchill

## Corresponding authors

Correspondence to [Gary A. Churchill](#) or [Steven P. Gygi](#).

## Additional information

**Supplementary information** is available in the online version of this Amendment.

## Supplementary information

### [Supplementary Table](#)

This file contains the corrected Supplementary Table 3 to the original paper.

### [Supplementary Table](#)

This file contains the corrected Supplementary Table 5 to the original paper.

### [Supplementary Table](#)

This file contains the corrected Supplementary Table 8 to the original paper.

## Rights and permissions

### [Reprints and Permissions](#)

## About this article

### Cite this article

Chick, J.M., Munger, S.C., Simecek, P. *et al.* Author Correction: Defining the consequences of genetic variation on a proteome-wide scale. *Nature* **606**, E16 (2022). <https://doi.org/10.1038/s41586-022-04920-w>

- Published: 07 June 2022
- Issue Date: 23 June 2022
- DOI: <https://doi.org/10.1038/s41586-022-04920-w>

### Share this article

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

[Get shareable link](#)

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

[Copy to clipboard](#)

Provided by the Springer Nature SharedIt content-sharing initiative

---

This article was downloaded by **calibre** from <https://www.nature.com/articles/s41586-022-04920-w>

- Publisher Correction
- [Published: 07 June 2022](#)

# Publisher Correction: Virtual communication curbs creative idea generation

- [Melanie S. Brucks](#) &
- [Jonathan Levav](#)

*Nature* volume **606**, page E17 (2022)

- 370 Accesses
- 1 Altmetric
- [Metrics details](#)

The [Original Article](#) was published on 27 April 2022

Correction to: *Nature* <https://doi.org/10.1038/s41586-022-04643-y>.  
Published online 27 April 2022.

In the version of this article initially published, the “Field experiment” section of the Methods was missing from the HTML version of the article and has now been restored

## Author information

### Authors and Affiliations

1. Marketing Division, Columbia Business School, Columbia University,  
New York, NY, USA

Melanie S. Brucks

2. Marketing Division, Stanford Graduate School of Business, Stanford  
University, Stanford, CA, USA

Jonathan Levav

## Corresponding author

Correspondence to [Melanie S. Brucks](#).

## Rights and permissions

[Reprints and Permissions](#)

## About this article

### Cite this article

Brucks, M.S., Levav, J. Publisher Correction: Virtual communication curbs creative idea generation. *Nature* **606**, E17 (2022).  
<https://doi.org/10.1038/s41586-022-04852-5>

- Published: 07 June 2022
- Issue Date: 23 June 2022
- DOI: <https://doi.org/10.1038/s41586-022-04852-5>

## Share this article

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

[Get shareable link](#)

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

[Copy to clipboard](#)

Provided by the Springer Nature SharedIt content-sharing initiative

---

This article was downloaded by **calibre** from <https://www.nature.com/articles/s41586-022-04852-5>

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

- Publisher Correction
- Open Access
- [Published: 07 June 2022](#)

# Publisher Correction: Genomic reconstruction of the SARS CoV-2 epidemic in England

- [Harald S. Vöhringer](#),
- [Theo Sanderson](#),
- [Matthew Sinnott](#),
- [Nicola De Maio](#),
- [Thuy Nguyen](#),
- [Richard Goater](#),
- [Frank Schwach](#),
- [Ian Harrison](#),
- [Joel Hellewell](#),
- [Cristina V. Ariani](#),
- [Sonia Gonçalves](#),
- [David K. Jackson](#),
- [Ian Johnston](#),
- [Alexander W. Jung](#),
- [Callum Saint](#),
- [John Sillitoe](#),
- [Maria Suciu](#),
- [Nick Goldman](#),
- [Jasmina Panovska-Griffiths](#),
- [The Wellcome Sanger Institute COVID-19 Surveillance Team](#),
- [The COVID-19 Genomics UK \(COG-UK\) Consortium\\*](#),
- [Ewan Birney](#),
- [Erik Volz](#),
- [Sebastian Funk](#),

- [Dominic Kwiatkowski](#),
- [Meera Chand](#),
- [Inigo Martincorena](#),
- [Jeffrey C. Barrett](#) &
- [Moritz Gerstung](#)

*Nature* volume **606**, page E18 (2022)

- 315 Accesses
- 2 Altmetric
- [Metrics details](#)

The [Original Article](#) was published on 14 October 2021

Correction to: *Nature* <https://doi.org/10.1038/s41586-021-04069-y>.

Published online 14 October 2021

In the version of this article initially published, the name of author Erik Volz appeared as “Erik M. Volz” in The COVID-19 Genomics UK (COG-UK) Consortium contributions listings. Further, the affiliation listed for COG-UK Consortium member Adam A. Witney was shown incorrectly (Imperial College London) and has been now amended to the Institute for Infection and Immunity, St George’s Hospital of London, London, UK. The changes have been made to the HTML and PDF versions of the article.

## Author information

### Authors and Affiliations

1. European Molecular Biology Laboratory, European Bioinformatics Institute EMBL-EBI, Hinxton, UK

Harald S. Vöhringer, Nicola De Maio, Alexander W. Jung, Nick Goldman, Ewan Birney & Moritz Gerstung

## 2. Wellcome Sanger Institute, Hinxton, UK

Theo Sanderson, Matthew Sinnott, Thuy Nguyen, Richard Goater, Frank Schwach, Cristina V. Ariani, Sonia Gonçalves, David K. Jackson, Ian Johnston, Callum Saint, John Sillitoe, Maria Suciu, Irina Abnizova, Louise Aigrain, Alex Alderton, Mozam Ali, Laura Allen, Roberto Amato, Ralph Anderson, Cristina Ariani, Siobhan Austin-Guest, Sendu Bala, Jeffrey Barrett, Andrew Bassett, Kristina Battleday, James Beal, Mathew Beale, Charlotte Beaver, Sam Bellany, Tristram Bellerby, Katie Bellis, Duncan Berger, Matt Berriman, Emma Betteridge, Paul Bevan, Simon Binley, Jason Bishop, Kirsty Blackburn, James Bonfield, Nick Boughton, Sam Bowker, Timothy Brendler-Spaeth, Iraad Bronner, Tanya Brooklyn, Sarah Kay Buddenborg, Robert Bush, Catarina Caetano, Alex Cagan, Nicola Carter, Joanna Cartwright, Tiago Carvalho Monteiro, Liz Chapman, Tracey-Jane Chillingworth, Peter Clapham, Richard Clark, Adrian Clarke, Catriona Clarke, Daryl Cole, Elizabeth Cook, Maria Coppola, Linda Cornell, Clare Cornwell, Craig Corton, Abby Crackett, Alison Cranage, Harriet Craven, Sarah Craw, Mark Crawford, Tim Cutts, Monika Dabrowska, Matt Davies, Robert Davies, Joseph Dawson, Callum Day, Aiden Densem, Thomas Dibling, Cat Dockree, David Dodd, Sunil Dogga, Matthew Dorman, Gordon Dougan, Martin Dougherty, Alexander Dove, Lucy Drummond, Eleanor Drury, Monika Dudek, Jillian Durham, Laura Durrant, Elizabeth Easthope, Sabine Eckert, Pete Ellis, Ben Farr, Michael Fenton, Marcella Ferrero, Neil Flack, Howerd Fordham, Grace Forsythe, Luke Foulser, Matt Francis, Audrey Fraser, Adam Freeman, Anastasia Galvin, Maria Garcia-Casado, Alex Gedny, Sophia Girgis, James Glover, Sonia Goncalves, Scott Goodwin, Oliver Gould, Marina Gourtovaia, Andy Gray, Emma Gray, Coline Griffiths, Yong Gu, Florence Guerin, Will Hamilton, Hannah Hanks, Ewan Harrison, Alexandria Harrott, Edward Harry, Julia Harvison, Paul Heath, Anastasia Hernandez-Koutoucheva, Rhiannon Hobbs, Dave Holland, Sarah Holmes, Gary Hornett, Nicholas Hough, Liz Huckle, Lena Hughes-Hallet, Adam Hunter, Stephen Inglis, Sameena Iqbal, Adam Jackson, David Jackson, Keith James, Dorota Jamrozy, Carlos Jimenez

Verdejo, Matthew Jones, Kalyan Kallepally, Leanne Kane, Keely Kay, Sally Kay, Jon Keatley, Alan Keith, Alison King, Lucy Kitchin, Matt Kleanthous, Martina Klimekova, Petra Korlevic, Ksenia Krasheninnikova, Greg Lane, Cordelia Langford, Adam Laverack, Katharine Law, Mara Lawniczak, Stefanie Lensing, Steven Leonard, Laura Letchford, Kevin Lewis, Amanah Lewis-Wade, Jennifer Liddle, Quan Lin, Sarah Lindsay, Sally Linsdell, Rich Livett, Stephanie Lo, Rhona Long, Jamie Lovell, Jon Lovell, Catherine Ludden, James Mack, Mark Maddison, Aleksei Makunin, Irfan Mamun, Jenny Mansfield, Neil Marriott, Matt Martin, Matthew Mayho, Shane McCarthy, Jo McClintock, Samantha McGuigan, Sandra McHugh, Liz McMinn, Carl Meadows, Emily Mobley, Robin Moll, Maria Morra, Leanne Morrow, Kathryn Murie, Sian Nash, Claire Nathwani, Plamena Naydenova, Alexandra Neaverson, Rachel Nelson, Ed Nerou, Jon Nicholson, Tabea Nimz, Guillaume G. Noell, Sarah O'Meara, Valeriu Ohan, Karen Oliver, Charles Olney, Doug Ormond, Agnes Oszlanczi, Steve Palmer, Yoke Fei Pang, Barbora Pardubska, Naomi Park, Aaron Parmar, Gaurang Patel, Minal Patel, Maggie Payne, Arabella Petersen, Deborah Plowman, Tom Preston, Liam Prestwood, Christoph Puethe, Michael Quail, Diana Rajan, Shavanti Rajatileka, Richard Rance, Suzannah Rawlings, Nicholas Redshaw, Joe Reynolds, Mark Reynolds, Simon Rice, Matt Richardson, Connor Roberts, Katrina Robinson, Melanie Robinson, David Robinson, Hazel Rogers, Eduardo Martin Rojo, Daljit Roopra, Mark Rose, Luke Rudd, Nicholas Salmon, David Saul, Frank Schwach, Carol Scott, Phil Seekings, Lesley Shirley, Alison Simms, Matthew Sinnott, Shanthi Sivadasan, Bart Siwek, Dale Sizer, Kenneth Skeldon, Jason Skelton, Joanna Slater-Tunstill, Lisa Sloper, Nathalie Smerdon, Chris Smith, Christen Smith, James Smith, Katie Smith, Michelle Smith, Sean Smith, Tina Smith, Leighton Sneade, Carmen Diaz Soria, Emily Souster, Andrew Sparkes, Michael Spencer-Chapman, Janet Squares, Robert Stanley, Claire Steed, Tim Stickland, Ian Still, Michael R. Stratton, Michelle Strickland, Allen Swann, Agnieszka Swiatkowska, Neil Sycamore, Emma Swift, Edward Symons, Suzanne Szluha, Emma Taluy, Nunu Tao, Katy Taylor, Sam Taylor, Stacey Thompson, Mark

Thompson, Mark Thomson, Nicholas Thomson, Scott Thurston, Gerry Tonkin-Hill, Dee Toombs, Benjamin Topping, Jaime Tovar-Corona, Daniel Ungureanu, James Uphill, Jana Urbanova, Philip Jansen Van Vuuren, Valerie Vancollie, Paul Voak, Danielle Walker, Matthew Walker, Matt Waller, Gary Ward, Charlie Weatherhogg, Niki Webb, Danni Weldon, Alan Wells, Eloise Wells, Luke Westwood, Theo Whipp, Thomas Whiteley, Georgia Whitton, Andrew Whitwham, Sara Widaa, Mia Williams, Mark Wilson, Sean Wright, Ewan M. Harrison, Sónia Gonçalves, David M. Aanensen, Dinesh Aggarwal, Anthony P. Underwood, Cordelia F. Langford, Dominic Kwiatkowski, Andrew R. Bassett, Michael H. Spencer Chapman, Theo Sanderson, Naomi R. Park, Robert M. Davies, Katherine L. Bellis, Matthew J. Dorman, Robin J. Moll, Jaime M. Tovar-Corona, Khalil Abudahab, Ben E. W. Taylor, Jon-Paul Keatley, Iraad F. Bronner, Ben W. Farr, Stefanie V. Lensing, Shane A. McCarthy, Michael A. Quail, Nicholas M. Redshaw, Scott A. J. Thurston, Jennifer Liddle, Dominic Kwiatkowski, Inigo Martincorena & Jeffrey C. Barrett

### 3. The Francis Crick Institute, London, UK

Theo Sanderson

### 4. Public Health England, London, UK

Frank Schwach, Ian Harrison, Catherine Ludden, Dinesh Aggarwal, Husam Osman, Meera Chand, Sharon J. Peacock, Peter Muir, Esther Robinson, Barry B. Vipond, Andrew Bosworth, Stephanie Hutchings, Hannah M. Pymont, Richard Myers, Alicia Thornton, Katerina Galai, Shazaad S. Y. Ahmad, Nicholas W. Machin, Richard Hopes, Chloe Bishop, Vicki Chalker, Elias Allara, Clare Pearson, David Bibby, Gavin Dabrera, Nicholas Ellaby, Eileen Gallagher, Jonathan Hubb, Angie Lackenby, David Lee, Nikos Manesis, Tamyo Mbisa, Steven Platt, Katherine A. Twohig & Meera Chand

### 5. London School of Hygiene & Tropical Medicine, London, UK

Joel Hellewell & Sebastian Funk

6. The Big Data Institute, Nuffield Department of Medicine, University of Oxford, Oxford, UK

Jasmina Panovska-Griffiths, Tanya Golubchik, David Bonsall, Katrina Lythgoe, Helen Fryer, Christophe Fraser & Laura Thomson

7. MRC Centre for Global Infectious Disease Analysis, Jameel Institute for Disease and Emergency Analytics, Imperial College London, London, UK

Erik Volz, Rob Johnson, Olivia Boyd, Lily Geidelberg, Manon Ragonnet-Cronin & Erik Volz

8. Guy's and St Thomas' NHS Foundation Trust, London, UK

Perminder Smith, Adela Alcolea, Natasha Ohemeng-Kumi, John Ramble, Jasveen Sehmi & Meera Chand

9. Division for AI in Oncology, German Cancer Research Centre DKFZ, Heidelberg, Germany

Moritz Gerstung

10. Department of Medicine, University of Cambridge, Cambridge, UK

Sharon Peacock, Ramin Sadri, Catarina Sousa, Catherine Ludden, M. Estee Torok, Dinesh Aggarwal, Alessandro M. Carabelli, Sharon J. Peacock, Beth Blane, MacGregor Cox, Patrick Maxwell, Ken Smith, Ellena Brooks, Carol M. Churcher, Mireille Fragakis, Katerina Galai, Andrew Jermy, Sarah Judges, Georgina M. McManus, Kim S. Smith, Elaine Westwick & Theresa Feltwell

11. Centre for Enzyme Innovation, University of Portsmouth, Portsmouth, UK

Samuel C. Robson, Angela H. Beckett & Ekaterina Shelest

12. School of Pharmacy & Biomedical Sciences, University of Portsmouth, Portsmouth, UK

Samuel C. Robson, Katie F. Loveson, Kate F. Cook, Christopher Fearn & Salman Goudarzi

13. Cardiff University, Cardiff, UK

Thomas R. Connor, Anna Price, Andrew Mack, Arthur Morriss, Thomas Whalley, Joel Southgate, Martyn Guest, Christine Kitchen, Ian Merrick, Robert Munn, Owen Jones, Catherine Bresner, William Fuller, Angela Marchbank & Trudy Workman

14. Public Health Wales, Cardiff, UK

Thomas R. Connor, Sally Corden, Catherine Moore, Joanne Watkins, Matthew Bull, Catryn Williams, Nicole Pacchiarini, Simon Cottrell, Sara Rey, David Heyburn, Chris Williams, Noel Craine, Malorie Perry, Alexander Adams, Tara Annett, Hibo Asad, Alec Birchley, Jason Coombes, Johnathan M. Evans, Laia Fina, Bree Gatica-Wilcox, Lauren Gilbert, Lee Graham, Jessica Hey, Ember Hilvers, Sophie Jones, Hannah Jones, Sara Kumziene-Summerhayes, Caoimhe McKerr, Jessica Powell, Georgia Pugh, Sarah Taylor, Joel Southgate, Alisha Davies, Elen De Lacy, Fatima Downing, Sue Edwards, Laura Gifford, Mari Morgan & Amy Gaskin

15. Institute of Microbiology and Infection, University of Birmingham, Birmingham, UK

Nicholas J. Loman, Sam Nicholls, Claire McMurray, Alan McNally, Joshua Quick, Radoslaw Poplawski, Joanne Stockton & Will Rowe

16. King's College London, London, UK

Rocio T. Martinez Nunez & Themoula Charalampous

17. University of Edinburgh, Edinburgh, UK

Andrew Rambaut, Kate E. Templeton, Áine O'Toole, Thomas Williams, Rachel Colquhoun, Verity Hill, Ben Jackson, J. T. McCrone, Nathan Medd, Emily Scher, Carlos E. Balcazar, Michael D. Gallagher, Daniel Maloney, Thomas D. Stanton & Kathleen A. Williamson

18. Centre for Clinical Infection and Diagnostics Research, Department of Infectious Diseases, Guy's and St Thomas' NHS Foundation Trust, London, UK

Luke B. Snell, Gaia Nebbia, Jonathan Edgeworth, Rahul Batra, Bindi Patel, Themoula Charalampous & Amita Patel

19. University College London Hospitals NHS Foundation Trust, London, UK

Eleni Nastouli, Catherine Houlihan, Matthew Byott, Sunando Roy, Judith Heaney, Judith Breuer, Moira J. Spyer, Rachel J. Williams, Charlotte A. Williams, John A. Hartley, Leah Ensell, Helen L. Lowe, Laurentiu Maftei, Matteo Mondani, Nadua Bayzid & Marius Cotic

20. Advanced Pathogen Diagnostics Unit, University College London Hospital, London, UK

Eleni Nastouli, Dan Frampton, Matthew Byott, Judith Heaney & Moira J. Spyer

21. Great Ormond Street Institute of Child Health, University College London, London, UK

Eleni Nastouli & Moira J. Spyer

22. Department of Infectious Diseases and Microbiology, Cambridge University Hospitals NHS Foundation Trust, Cambridge, UK

M. Estee Torok & William L. Hamilton

23. Division of Virology, Department of Pathology, University of Cambridge, Cambridge, UK

Ian G. Goodfellow, Yasmin Chaudhry, Rhys Izuagbe, Aminu S. Jahun, Iliana Georgana, Myra Hosmillo & Malte L. Pinckert

24. University Hospital Southampton NHS Foundation Trust, Southampton, UK

Jacqui A. Prieto, Kordo Saeed, Stephen Aplin, Matthew Harvey, Thea Sass, Helen Umpleby, Helen Wheeler, Emanuela Pelosi, Siona Silviera, Eleri Wilson-Davies, Adhyana I. K. Mahanama, Buddhini Samaraweera & Sarah Jeremiah

25. School of Health Sciences, University of Southampton, Southampton, UK

Jacqui A. Prieto

26. School of Medicine, University of Southampton, Southampton, UK

Kordo Saeed

27. Division of Infection and Immunity, University College London, London, UK

Catherine Houlihan & Dan Frampton

28. University of Brighton, Brighton, UK

Giselda Bucca, Colin P. Smith & Andrew R. Hesketh

29. Institute for Infection and Immunity, St George's University of London, London, UK

Adam A. Witney, Cassie F. Pope, Kenneth G. Laing & Irene M. Monahan

30. Infection Care Group, St George's University Hospitals NHS Foundation Trust, London, UK

Cassie F. Pope

31. MRC–University of Glasgow Centre for Virus Research, Glasgow, UK

Emma C. Thomson, James G. Shepherd, David L. Robertson, William T. Harvey, Joseph Hughes, Richard J. Orton, Ana da Silva Filipe, Igor Starinskij, Sreenu Vattipally, Afrida Mukaddas, Derek W. Wright, Alice Broos, Daniel Mair, Jenna Nichols, Kyriaki Nomikou, Lily Tong & Ioulia Tsatsani

32. University of Cambridge, Cambridge, UK

Ewan M. Harrison, Ravi K. Gupta, Michael H. Spencer Chapman, Chris Ruis, Sophia T. Girgis, Katherine L. Bellis, Leanne M. Kermack, Claire Cormie, Joana Dias, Sally Forrest, Ellen E. Higginson, Mailis Maes, Jamie Young, Elias Allara, Clare Pearson & Shane A. McCarthy

33. Queen's University Belfast, Belfast, UK

Fiona Rogan, Derek J. Fairley, David A. Simpson, Marc Fuchs, Julia Miskelly, Stephen Bridgett, Timofey Skvortsov, Declan T. Bradley, Zoltan Molnar, Chris Baxter, Sílvia F. Carvalho, Deborah Lavin, Arun Mariappan, Clara Radulescu, Aditi Singh & Miao Tang

34. Blackpool Teaching Hospitals NHS Foundation Trust, Blackpool, UK

Shaun M. Beckwith, Abigail Murray & Dawn Singleton

35. Hull University Teaching Hospitals NHS Trust, Hull, UK

Kirstine Eastick, Patrick J. Lillie, Phillipa J. Burns, Kavitha Gajee, Katie Kitchman & William Everson

36. University Hospitals Dorset NHS Foundation Trust, Poole, UK

Liz A. Sheridan, Dorian Crudgington & Ben Macklin

37. University Hospitals Sussex NHS Foundation Trust, Worthing, UK

Paul Randell, Mohammed O. Hassan-Ibrahim, Michelle J. Erkiet, Nicola J. Chaloner, Benjamin J. Cogger, Lisa J. Easton, Hannah Huckson, Jonathan Lewis, Sarah Lowdon, Cassandra S. Malone, Florence Munemo, Manasa Mutingwende, Roberto Nicodemi, Olga Podplomyk & Thomas Somassa

38. University of Exeter, Exeter, UK

Leigh M. Jackson, Ben Temperton, Stephen L. Michell, Aaron R. Jeffries, Christopher R. Jones, Bridget A. Knight, Robin Manley, Michelle L. Michelsen, Christine M. Sambles, David J. Studholme & Joanna Warwick-Dugdale

39. Belfast Health & Social Care Trust, Belfast, UK

Derek J. Fairley, James P. McKenna & Tanya Curran

40. Deep Seq, School of Life Sciences, Queens Medical Centre, University of Nottingham, Nottingham, UK

Matthew W. Loose, Matthew Carlile, Nadine Holmes, Christopher Moore, Victoria Wright, Johnny Debebe & Fei Sang

41. East Kent Hospitals University NHS Foundation Trust, Canterbury, UK

Samuel Moses & Hannah Lowe

42. University of Kent, Canterbury, UK

Samuel Moses

43. Hub for Biotechnology in the Built Environment, Northumbria University, Newcastle upon Tyne, UK

Darren L. Smith, Matthew Bashton & Gregory R. Young

44. Northumbria University, Newcastle upon Tyne, UK

Darren L. Smith, Matthew Bashton, Gregory R. Young, Clare M. McCann, Andrew Nelson, Matthew R. Crown, John H. Henderson, Amy Hollis, William Stanley & Wen C. Yew

45. NU-OMICS, Northumbria University, Newcastle upon Tyne, UK

Darren L. Smith, Clare M. McCann & Andrew Nelson

46. Centre for Genomic Pathogen Surveillance, University of Oxford, Oxford, UK

David M. Aanensen, Anthony P. Underwood, Khalil Abudahab, Mirko Menegazzo, Ben E. W. Taylor & Corin A. Yeats

47. Public Health England, Cambridge, Cambridge, UK

Martin D. Curran & Surendra Parmar

48. University of Sheffield, Sheffield, UK

Matthew D. Parker, Thushan I. de Silva, Dennis Wang, Mehmet Yavus, Samantha E. Hansford, Benjamin H. Foulkes, Marta Gallis, Hailey R. Hornsby, Stavroula F. Louka, Manoj Pohare, Paige Wolverson, Peijun Zhang, Timothy M. Freeman, Sharon N. Hsu, Benjamin B. Lindsey, Nikki Smith, Cariad Evans, Kate Johnson, David G. Partridge, Mohammad Raza, Alexander J. Keeley, Adrienn Angyal, Luke R. Green & Max Whiteley

49. Portsmouth Hospitals University NHS Trust, Portsmouth, UK

Sharon Glaysher, Scott Elliott, Kelly Bicknell, Robert Impey & Sarah Wyllie

50. NHS Lothian, Edinburgh, UK

Kate E. Templeton, Rebecca Dewar, Martin P. McHugh, Elizabeth Wastnedge, Seb Cotton & Abbie Gallagher

51. NHS Greater Glasgow and Clyde, Glasgow, UK

Rory N. Gunson, Rachel Blacow & Jon Perkins

52. Quadram Institute Bioscience, Norwich, UK

Justin O'Grady, Andrew J. Page, Gemma L. Kay, Nabil-Fareed Alikhan, Alexander J. Trotter, Leonardo de Oliveira Martins, Alison E. Mather, Lizzie Meadows, Alp Aydin, David J. Baker, Ebenezer Foster-Nyarko, Sophie J. Prosolek, Steven Rudder & Thanh Le-Viet

53. University of East Anglia, Norwich, UK

Justin O'Grady & Rose K. Davidson

54. University of Oxford, Oxford, UK

Dominic Kwiatkowski

55. Hampshire Hospitals NHS Foundation Trust, Basingstoke, UK

Nicholas Cortes, Nathan Moore, Claire Thomas, Stephen P. Kidd, Beatrice Bertolusso, Jessica Lynch, Gabrielle Vernet & Rebecca Williams

56. Royal Free London NHS Foundation Trust, London, UK

Tabitha W. Mahungu, Tanzina Haque, Jennifer Hart, Dianne Irish-Tavares & Eric Witele

57. South Tees Hospitals NHS Foundation Trust, Middlesbrough, UK

Steven Liggett, Craig Mower, Louisa K. Watson, Paul Baker, Stephen Bonner, Sarah Essex & Leanne J. Murray

58. School of Biological Sciences, University of Portsmouth, Portsmouth, UK

Angela H. Beckett, Yann Bourgeois & Ekaterina Shelest

59. Public Health Scotland, Edinburgh, UK

Matthew T. G. Holden, Sharif Shaaban, Stefan Rooke & Gonzalo Yebra

60. Health Services Laboratories, London, UK

Lisa J. Levett, Judith Heaney, Paul R. Grant, Stuart Kirk, Wendy Chatterton & Monika Pusok

61. Heartlands Hospital, Birmingham, UK

Husam Osman, Esther Robinson, Li Xu-McCrae & Andrew Bosworth

62. University of Liverpool, Liverpool, UK

Alistair C. Darby, Steve Paterson, Miren Ituriza-Gomara, Anita O. Lucaci, Sam T. Haldenby, Kathryn A. Jackson, Lance Turtle, Richard Eccles, Matthew Gemmell, Richard Gregory, Margaret Hughes, Charlotte Nelson, Lucille Rainbow, Edith E. Vamos, Hermione J. Webster, Mark Whitehead & Claudia Wierzbicki

63. Department of Zoology, University of Oxford, Oxford, UK

Oliver G. Pybus, Louis du Plessis, Moritz U. G. Kraemer, Jayna Raghwani, Alex E. Zarebski, Stephen W. Attwood, Marina Escalera Zamudio, Sarah Francois, Bernardo Gutierrez & Tetyana I. Vasylyeva

64. MRC Biostatistics Unit, University of Cambridge, Cambridge, UK

Daniela de Angelis, Chris J. Illingworth, Chris Jackson & David Pascall

65. Microbiology Department, Buckinghamshire Healthcare NHS Trust, Aylesbury, UK

Nick Wong, David T. Pritchard, Debbie Binns & Victoria James

66. The Newcastle upon Tyne Hospitals NHS Foundation Trust, Newcastle upon Tyne, UK

Yusri Taha, Jennifer Collins, Gary Eltringham, Shirelle Burton-Fanning, Brendan A. I. Payne & Sheila Waugh

67. University of St Andrews, St Andrews, UK

Martin P. McHugh

68. Imperial College Healthcare NHS Trust, London, UK

Siddharth Mookerjee, Alison Cox, Pinglawathee Madona, Marcus Pond, Paul A. Randell, Alison H. Holmes, James R. Price & Frances Bolt

69. NIHR Health Protection Research Unit in HCAI and AMR, Imperial College London, London, UK

Siddharth Mookerjee, Alison H. Holmes, James R. Price & Frances Bolt

70. Cambridge University Hospitals NHS Foundation Trust, Cambridge, UK

Ben Warne

71. Department of Microbiology, South West London Pathology, London, UK

Joshua F. Taylor & Ngee Keong Tan

72. Betsi Cadwaladr University Health Board, Wrexham, UK

Helen Adams

73. Institute of Biodiversity, Animal Health & Comparative Medicine,  
Glasgow, UK

William T. Harvey & Katherine L. Smollett

74. Norfolk County Council, Norfolk, UK

Lewis G. Spurgin & Louise Smith

75. Great Ormond Street Hospital for Children NHS Foundation Trust,  
London, UK

Judith Breuer, Julianne R. Brown, Kathryn A. Harris, Nathaniel  
Storey, Laura Atkinson, Jack C. D. Lee & Divya Shah

76. Wellcome Centre for Human Genetics, Nuffield Department of  
Medicine, University of Oxford, Oxford, UK

John A. Todd, David Buck, Angie Green, George MacIntyre-  
Cockett & Amy Trebes

77. Cardiff and Vale University Health Board, Cardiff, UK

Michaela John, Sian Morgan, Safiah Afifi, Robert Beer, Joshua  
Maksimovic, Kathryn McCluggage Masters & Karla Spellman

78. Turnkey Laboratory, University of Birmingham, Birmingham, UK

Alan McNally, Fiona Ashford, Angus Best, Liam Crawford, Nicola  
Cumley, Megan Mayhew, Oliver Mogram, Jeremy Mirza, Emma  
Moles-Garcia & Benita Percival

79. Norfolk and Norwich University Hospitals NHS Foundation Trust,  
Norwich, UK

Samir Dervisevic, Rachael Stanley & Lindsay Coupland

80. Royal Brompton and Harefield Hospitals, London, UK

Newara A. Ramadan

81. The Queen Elizabeth Hospital King's Lynn NHS Foundation Trust,  
King's Lynn, UK

Christopher Jeanes

82. Whittington Health NHS Trust, London, UK

Jana Catalan & Neil Jones

83. Barking, Havering and Redbridge University Hospitals NHS Trust,  
London, UK

Amy Ash & Cherian Koshy

84. Bournemouth University, Bournemouth, UK

Magdalena Barrow, Sarah L. Buchan & Anna Mantzouratou

85. Clinical Microbiology Department, Queens Medical Centre,  
Nottingham University Hospitals NHS Trust, Nottingham, UK

Gemma Clark, Louise Berry, Tim Boswell, Vicki M. Fleming, Hannah C. Howson-Wells, Amelia Joseph, Manjinder Khakh & Michelle M. Lister

86. Clinical Microbiology, University Hospitals of Leicester NHS Trust,  
Leicester, UK

Christopher W. Holmes, Paul W. Bird, Karlie Fallon, Thomas Helmer, Claire L. McMurray, Mina Odedra, Jessica Shaw, Julian W. Tang & Nicholas J. Willford

87. County Durham and Darlington NHS Foundation Trust, Durham, UK

Sharon Campbell, Victoria Blakey, Veena Raviprakash, Nicola Sheriff & Lesley-Anne Williams

88. Department of Microbiology, Kettering General Hospital, Kettering, UK

Thomas Davis, Sahar Eldirdiri & Anita Kenyon

89. Barts Health NHS Trust, London, UK

Kathryn A. Harris

90. North West London Pathology, London, UK

Alison Cox, Pinglawathee Madona, Marcus Pond & Paul A. Randell

91. Maidstone and Tunbridge Wells NHS Trust, Maidstone, UK

Karen T. Withell & Graciela Sluga

92. Microbiology, Royal Oldham Hospital, Oldham, UK

Cheryl Williams

93. North Cumbria Integrated Care NHS Foundation Trust, Carlisle, UK

Clive Graham, Edward Barton, Debra Padgett & Garren Scott

94. North Tees and Hartlepool NHS Foundation Trust, Stockton on Tees, UK

Rebecca Denton-Smith, Emma Swindells, Robyn Turnbull & Jane Greenaway

95. Path Links, Northern Lincolnshire and Goole NHS Foundation Trust, Grimsby, UK

Tim J. Sloan, Phillip Clarke, Nichola Duckworth & Sarah Walsh

96. Queen Elizabeth Hospital, Birmingham, UK

Anna Casey & Liz Ratcliffe

97. Royal Devon and Exeter NHS Foundation Trust, Exeter, UK

Christopher R. Jones, Bridget A. Knight, Jennifer Poyner, Cressida Auckland & Helen Morcrette

98. Virology, School of Life Sciences, Queens Medical Centre, University of Nottingham, Nottingham, UK

Patrick C. McClure, Jonathan Ball, Timothy Byaruhanga, Joseph G. Chappell, Jayasree Dey, Jack D. Hill & Emily J. Park

99. Swansea University, Swansea, UK

Ronan A. Lyons

100. Viapath, Guy's and St Thomas' NHS Foundation Trust, and King's College Hospital NHS Foundation Trust, London, UK

Perminder Smith, Adela Alcolea, Natasha Ohemeng-Kumi, John Ramble & Jasveen Sehmi

101. Sheffield Teaching Hospitals NHS Foundation Trust, Sheffield, UK

Mehmet Yavus, Cariad Evans, Kate Johnson, David G. Partridge & Mohammad Raza

102. Cambridge Stem Cell Institute, University of Cambridge, Cambridge, UK

Nicola Reynolds

103. West of Scotland Specialist Virology Centre, NHS Greater Glasgow and Clyde, Glasgow, UK

Lynne Ferguson, Emily J. Goldstein, Alasdair Maclean & Rachael Tomb

104. Public Health Agency, Belfast, UK

Tim Wyatt & Declan T. Bradley

105. Northumbria Healthcare NHS Foundation Trust, Newcastle upon Tyne, UK

Giles Idle & Kevin Cole

106. University of Southampton, Southampton, UK

Matilde Mori

107. East Suffolk and North Essex NHS Foundation Trust, Colchester, UK

Luke Bedford

108. East Sussex Healthcare NHS Trust, St Leonards-on-Sea, UK

James S. Cargill & Warwick Hughes

109. Gateshead Health NHS Foundation Trust, Gateshead, UK

Jonathan Moore & Susanne Stonehouse

110. Isle of Wight NHS Trust, Newport, UK

Anibolina Castigador & Emily Macnaughton

111. King's College Hospital NHS Foundation Trust, London, UK

Kate El Bouzidi, Temi Lampejo & Malur Sudhanva

112. Liverpool Clinical Laboratories, Liverpool, UK

Cassie Breen

113. Manchester University NHS Foundation Trust, Manchester, UK

Shazaad S. Y. Ahmad, Ryan P. George & Nicholas W. Machin

114. North Middlesex University Hospital NHS Trust, London, UK

Aidan Cross & Mariyam Mirfenderesky

115. Southwest Pathology Services, Taunton, UK

Andrew I. Lawton

116. The Royal Marsden NHS Foundation Trust, London, UK

Andrea N. Gomes, Maimuna Kimuli & Darren R. Murray

117. The Royal Wolverhampton NHS Trust, Wolverhampton, UK

Paula Ashfield & Donald Dobie

118. University of Birmingham, Birmingham, UK

Andrew Beggs & Alex Richter

119. Watford General Hospital, Watford, UK

Arezou Fanaie, Rachel A. Hilson & Geraldine Yaze

120. Guy's and St Thomas' Biomedical Research Centre, London, UK

Flavia Flaviani

121. Newcastle University, Newcastle, UK

Sarah O'Brien, Steven Rushton & Roy Sanderson

## **Consortia**

**The Wellcome Sanger Institute COVID-19 Surveillance Team**

- Irina Abnizova
- , Louise Aigrain
- , Alex Alderton
- , Mozam Ali
- , Laura Allen
- , Roberto Amato
- , Ralph Anderson
- , Cristina Ariani
- , Siobhan Austin-Guest
- , Sendu Bala
- , Jeffrey Barrett
- , Andrew Bassett
- , Kristina Battleday
- , James Beal
- , Mathew Beale
- , Charlotte Beaver
- , Sam Bellany
- , Tristram Bellerby
- , Katie Bellis
- , Duncan Berger
- , Matt Berriman
- , Emma Betteridge
- , Paul Bevan
- , Simon Binley
- , Jason Bishop
- , Kirsty Blackburn
- , James Bonfield
- , Nick Boughton
- , Sam Bowker
- , Timothy Brendler-Spaeth
- , Iraad Bronner
- , Tanya Brooklyn
- , Sarah Kay Buddenborg
- , Robert Bush
- , Catarina Caetano
- , Alex Cagan
- , Nicola Carter

- , Joanna Cartwright
- , Tiago Carvalho Monteiro
- , Liz Chapman
- , Tracey-Jane Chillingworth
- , Peter Clapham
- , Richard Clark
- , Adrian Clarke
- , Catriona Clarke
- , Daryl Cole
- , Elizabeth Cook
- , Maria Coppola
- , Linda Cornell
- , Clare Cornwell
- , Craig Corton
- , Abby Crackett
- , Alison Cranage
- , Harriet Craven
- , Sarah Craw
- , Mark Crawford
- , Tim Cutts
- , Monika Dabrowska
- , Matt Davies
- , Robert Davies
- , Joseph Dawson
- , Callum Day
- , Aiden Densem
- , Thomas Dibling
- , Cat Dockree
- , David Dodd
- , Sunil Dogga
- , Matthew Dorman
- , Gordon Dougan
- , Martin Dougherty
- , Alexander Dove
- , Lucy Drummond
- , Eleanor Drury
- , Monika Dudek

- , Jillian Durham
- , Laura Durrant
- , Elizabeth Easthope
- , Sabine Eckert
- , Pete Ellis
- , Ben Farr
- , Michael Fenton
- , Marcella Ferrero
- , Neil Flack
- , Howerd Fordham
- , Grace Forsythe
- , Luke Foulser
- , Matt Francis
- , Audrey Fraser
- , Adam Freeman
- , Anastasia Galvin
- , Maria Garcia-Casado
- , Alex Gedny
- , Sophia Girgis
- , James Glover
- , Sonia Goncalves
- , Scott Goodwin
- , Oliver Gould
- , Marina Gourtovaia
- , Andy Gray
- , Emma Gray
- , Coline Griffiths
- , Yong Gu
- , Florence Guerin
- , Will Hamilton
- , Hannah Hanks
- , Ewan Harrison
- , Alexandria Harrott
- , Edward Harry
- , Julia Harvison
- , Paul Heath
- , Anastasia Hernandez-Koutoucheva

- , Rhiannon Hobbs
- , Dave Holland
- , Sarah Holmes
- , Gary Hornett
- , Nicholas Hough
- , Liz Huckle
- , Lena Hughes-Hallet
- , Adam Hunter
- , Stephen Inglis
- , Sameena Iqbal
- , Adam Jackson
- , David Jackson
- , Keith James
- , Dorota Jamrozy
- , Carlos Jimenez Verdejo
- , Ian Johnston
- , Matthew Jones
- , Kalyan Kallepally
- , Leanne Kane
- , Keely Kay
- , Sally Kay
- , Jon Keatley
- , Alan Keith
- , Alison King
- , Lucy Kitchin
- , Matt Kleanthous
- , Martina Klimekova
- , Petra Korlevic
- , Ksenia Krasheninnkova
- , Dominic Kwiatkowski
- , Greg Lane
- , Cordelia Langford
- , Adam Laverack
- , Katharine Law
- , Mara Lawniczak
- , Stefanie Lensing
- , Steven Leonard

- , Laura Letchford
- , Kevin Lewis
- , Amanah Lewis-Wade
- , Jennifer Liddle
- , Quan Lin
- , Sarah Lindsay
- , Sally Linsdell
- , Rich Livett
- , Stephanie Lo
- , Rhona Long
- , Jamie Lovell
- , Jon Lovell
- , Catherine Ludden
- , James Mack
- , Mark Maddison
- , Aleksei Makunin
- , Irfan Mamun
- , Jenny Mansfield
- , Neil Marriott
- , Matt Martin
- , Inigo Martincorena
- , Matthew Mayho
- , Shane McCarthy
- , Jo McClintonck
- , Samantha McGuigan
- , Sandra McHugh
- , Liz McMinn
- , Carl Meadows
- , Emily Mobley
- , Robin Moll
- , Maria Morra
- , Leanne Morrow
- , Kathryn Murie
- , Sian Nash
- , Claire Nathwani
- , Plamena Naydenova
- , Alexandra Neaverson

- , Rachel Nelson
- , Ed Nerou
- , Jon Nicholson
- , Tabea Nimz
- , Guillaume G. Noell
- , Sarah O'Meara
- , Valeriu Ohan
- , Karen Oliver
- , Charles Olney
- , Doug Ormond
- , Agnes Oszlanczi
- , Steve Palmer
- , Yoke Fei Pang
- , Barbora Pardubska
- , Naomi Park
- , Aaron Parmar
- , Gaurang Patel
- , Minal Patel
- , Maggie Payne
- , Sharon Peacock
- , Arabella Petersen
- , Deborah Plowman
- , Tom Preston
- , Liam Prestwood
- , Christoph Puethe
- , Michael Quail
- , Diana Rajan
- , Shavanthi Rajatileka
- , Richard Rance
- , Suzannah Rawlings
- , Nicholas Redshaw
- , Joe Reynolds
- , Mark Reynolds
- , Simon Rice
- , Matt Richardson
- , Connor Roberts
- , Katrina Robinson

- , Melanie Robinson
- , David Robinson
- , Hazel Rogers
- , Eduardo Martin Rojo
- , Daljit Roopra
- , Mark Rose
- , Luke Rudd
- , Ramin Sadri
- , Nicholas Salmon
- , David Saul
- , Frank Schwach
- , Carol Scott
- , Phil Seekings
- , Lesley Shirley
- , John Sillitoe
- , Alison Simms
- , Matthew Sinnott
- , Shanthi Sivadasan
- , Bart Siwek
- , Dale Sizer
- , Kenneth Skeldon
- , Jason Skelton
- , Joanna Slater-Tunstill
- , Lisa Sloper
- , Nathalie Smerdon
- , Chris Smith
- , Christen Smith
- , James Smith
- , Katie Smith
- , Michelle Smith
- , Sean Smith
- , Tina Smith
- , Leighton Sneade
- , Carmen Diaz Soria
- , Catarina Sousa
- , Emily Souster
- , Andrew Sparkes

- , Michael Spencer-Chapman
- , Janet Squares
- , Robert Stanley
- , Claire Steed
- , Tim Stickland
- , Ian Still
- , Michael R. Stratton
- , Michelle Strickland
- , Allen Swann
- , Agnieszka Swiatkowska
- , Neil Sycamore
- , Emma Swift
- , Edward Symons
- , Suzanne Szluha
- , Emma Taluy
- , Nunu Tao
- , Katy Taylor
- , Sam Taylor
- , Stacey Thompson
- , Mark Thompson
- , Mark Thomson
- , Nicholas Thomson
- , Scott Thurston
- , Gerry Tonkin-Hill
- , Dee Toombs
- , Benjamin Topping
- , Jaime Tovar-Corona
- , Daniel Ungureanu
- , James Uphill
- , Jana Urbanova
- , Philip Jansen Van Vuuren
- , Valerie Vancollie
- , Paul Voak
- , Danielle Walker
- , Matthew Walker
- , Matt Waller
- , Gary Ward

- , Charlie Weatherhogg
- , Niki Webb
- , Danni Weldon
- , Alan Wells
- , Eloise Wells
- , Luke Westwood
- , Theo Whipp
- , Thomas Whiteley
- , Georgia Whitton
- , Andrew Whitwham
- , Sara Widaa
- , Mia Williams
- , Mark Wilson
- & Sean Wright

## The COVID-19 Genomics UK (COG-UK) Consortium\*

- **Funding acquisition, leadership and supervision, metadata curation, project administration, samples and logistics, sequencing and analysis, software and analysis tools, and visualization**
  - Samuel C. Robson
- **Funding acquisition, leadership and supervision, metadata curation, project administration, samples and logistics, sequencing and analysis, and software and analysis tools**
  - Thomas R. Connor
  - & Nicholas J. Loman
- **Leadership and supervision, metadata curation, project administration, samples and logistics, sequencing and analysis, software and analysis tools, and visualization**
  - Tanya Golubchik

- **Funding acquisition, leadership and supervision, metadata curation, samples and logistics, sequencing and analysis, and visualization**
  - Rocio T. Martinez Nunez
- **Funding acquisition, leadership and supervision, project administration, samples and logistics, sequencing and analysis, and software and analysis tools**
  - David Bonsall
- **Funding acquisition, leadership and supervision, project administration, sequencing and analysis, software and analysis tools, and visualization**
  - Andrew Rambaut
- **Funding acquisition, metadata curation, project administration, samples and logistics, sequencing and analysis, and software and analysis tools**
  - Luke B. Snell
- **Leadership and supervision, metadata curation, project administration, samples and logistics, software and analysis tools, and visualization**
  - Rich Livett
- **Funding acquisition, leadership and supervision, metadata curation, project administration, and samples and logistics**
  - Catherine Ludden

- **Funding acquisition, leadership and supervision, metadata curation, samples and logistics, and sequencing and analysis**
  - Sally Corden
  - & Eleni Nastouli
- **Funding acquisition, leadership and supervision, metadata curation, sequencing and analysis, and software and analysis tools**
  - Gaia Nebbia
- **Funding acquisition, leadership and supervision, project administration, samples and logistics, and sequencing and analysis**
  - Ian Johnston
- **Leadership and supervision, metadata curation, project administration, samples and logistics, and sequencing and analysis**
  - Katrina Lythgoe
  - , M. Estee Torok
  - & Ian G. Goodfellow
- **Leadership and supervision, metadata curation, project administration, samples and logistics, and visualization**
  - Jacqui A. Prieto
  - & Kordo Saeed
- **Leadership and supervision, metadata curation, project administration, sequencing and analysis, and software and analysis tools**
  - David K. Jackson

- **Leadership and supervision, metadata curation, samples and logistics, sequencing and analysis, and visualization**
  - Catherine Houlihan
- **Leadership and supervision, metadata curation, sequencing and analysis, software and analysis tools, and visualization**
  - Dan Frampton
- **Metadata curation, project administration, samples and logistics, sequencing and analysis, and software and analysis tools**
  - William L. Hamilton
  - & Adam A. Witney
- **Funding acquisition, samples and logistics, sequencing and analysis, and visualization**
  - Giselda Bucca
- **Funding acquisition, leadership and supervision, metadata curation and project administration**
  - Cassie F. Pope
- **Funding acquisition, leadership and supervision, metadata curation, and samples and logistics**
  - Catherine Moore
- **Funding acquisition, leadership and supervision, metadata curation, and sequencing and analysis**
  - Emma C. Thomson

- **Funding acquisition, leadership and supervision, project administration, and samples and logistics**
  - Ewan M. Harrison
- **Funding acquisition, leadership and supervision, sequencing and analysis, and visualization**
  - Colin P. Smith
- **Leadership and supervision, metadata curation, project administration, and sequencing and analysis**
  - Fiona Rogan
- **Leadership and supervision, metadata curation, project administration, and samples and logistics**
  - Shaun M. Beckwith
  - , Abigail Murray
  - , Dawn Singleton
  - , Kirstine Eastick
  - , Liz A. Sheridan
  - , Paul Randell
  - , Leigh M. Jackson
  - , Cristina V. Ariani
  - & Sónia Gonçalves
- **Leadership and supervision, metadata curation, samples and logistics, and sequencing and analysis**
  - Derek J. Fairley
  - , Matthew W. Loose
  - & Joanne Watkins

- **Leadership and supervision, metadata curation, samples and logistics, and visualization**
  - Samuel Moses
- **Leadership and supervision, metadata curation, sequencing and analysis, and software and analysis tools**
  - Sam Nicholls
  - , Matthew Bull
  - & Roberto Amato
- **Leadership and supervision, project administration, samples and logistics, and sequencing and analysis**
  - Darren L. Smith
- **Leadership and supervision, sequencing and analysis, software and analysis tools, and visualization**
  - David M. Aanensen
  - & Jeffrey C. Barrett
- **Metadata curation, project administration, samples and logistics, and sequencing and analysis**
  - Dinesh Aggarwal
  - , James G. Shepherd
  - , Martin D. Curran
  - & Surendra Parmar
- **Metadata curation, project administration, sequencing and analysis, and software and analysis tools**
  - Matthew D. Parker

- **Metadata curation, samples and logistics, sequencing and analysis, and software and analysis tools**
  - Catryn Williams
- **Metadata curation, samples and logistics, sequencing and analysis, and visualization**
  - Sharon Glaysher
- **Metadata curation, sequencing and analysis, software and analysis tools, and visualization**
  - Anthony P. Underwood
  - , Matthew Bashton
  - , Nicole Pacchiarini
  - , Katie F. Loveson
  - & Matthew Byott
- **Project administration, sequencing and analysis, software and analysis tools, and visualization**
  - Alessandro M. Carabelli
- **Funding acquisition, leadership and supervision, and metadata curation**
  - Kate E. Templeton
- **Funding acquisition, leadership and supervision, and project administration**
  - Thushan I. de Silva
  - , Dennis Wang
  - , Cordelia F. Langford

- & John Sillitoe
- **Funding acquisition, leadership and supervision, and samples and logistics**
  - Rory N. Gunson
- **Funding acquisition, leadership and supervision, and sequencing and analysis**
  - Simon Cottrell
  - , Justin O’Grady
  - & Dominic Kwiatkowski
- **Leadership and supervision, metadata curation and project administration**
  - Patrick J. Lillie
- **Leadership and supervision, metadata curation, and samples and logistics**
  - Nicholas Cortes
  - , Nathan Moore
  - , Claire Thomas
  - , Phillipa J. Burns
  - , Tabitha W. Mahungu
  - & Steven Liggett
- **Leadership and supervision, metadata curation, and sequencing and analysis**
  - Angela H. Beckett
  - & Matthew T. G. Holden

- **Leadership and supervision, project administration, and samples and logistics**
  - Lisa J. Levett
  - , Husam Osman
  - & Mohammed O. Hassan-Ibrahim
- **Leadership and supervision, project administration, and sequencing and analysis**
  - David A. Simpson
- **Leadership and supervision, samples and logistics, and sequencing and analysis**
  - Meera Chand
  - , Ravi K. Gupta
  - , Alistair C. Darby
  - & Steve Paterson
- **Leadership and supervision, sequencing and analysis, and software and analysis tools**
  - Oliver G. Pybus
  - , Erik Volz
  - , Daniela de Angelis
  - , David L. Robertson
  - , Andrew J. Page
  - & Inigo Martincorena
- **Leadership and supervision, sequencing and analysis, and visualization**
  - Louise Aigrain
  - & Andrew R. Bassett

- **Metadata curation, project administration, and samples and logistics**
  - Nick Wong
  - , Yusri Taha
  - , Michelle J. Erkier
  - & Michael H. Spencer Chapman
- **Metadata curation, project administration, and sequencing and analysis**
  - Rebecca Dewar
  - & Martin P. McHugh
- **Metadata curation, project administration, and software and analysis tools**
  - Siddharth Mookerjee
- **Metadata curation, project administration and visualization**
  - Stephen Aplin
  - , Matthew Harvey
  - , Thea Sass
  - , Helen Umpleby
  - & Helen Wheeler
- **Metadata curation, samples and logistics, and sequencing and analysis**
  - James P. McKenna
  - , Ben Warne
  - , Joshua F. Taylor
  - , Yasmin Chaudhry
  - , Rhys Izuagbe
  - , Aminu S. Jahun

- , Gregory R. Young
  - , Claire McMurray
  - , Clare M. McCann
  - , Andrew Nelson
  - & Scott Elliott
- **Metadata curation, samples and logistics, and visualization**
    - Hannah Lowe
  - **Metadata curation, sequencing and analysis, and software and analysis tools**
    - Anna Price
    - , Matthew R. Crown
    - , Sara Rey
    - , Sunando Roy
    - & Ben Temperton
  - **Metadata curation, sequencing and analysis, and visualization**
    - Sharif Shaaban
    - & Andrew R. Hesketh
  - **Project administration, samples and logistics, and sequencing and analysis**
    - Kenneth G. Laing
    - , Irene M. Monahan
    - & Judith Heaney
  - **Project administration, samples and logistics, and visualization**
    - Emanuela Pelosi
    - , Siona Silviera
    - & Eleri Wilson-Davies

- **Samples and logistics, software and analysis tools, and visualization**
  - Helen Fryer
- **Sequencing and analysis, software and analysis tools, and visualization**
  - Helen Adams
  - , Louis du Plessis
  - , Rob Johnson
  - , William T. Harvey
  - , Joseph Hughes
  - , Richard J. Orton
  - , Lewis G. Spurgin
  - , Yann Bourgeois
  - , Chris Ruis
  - , Áine O'Toole
  - , Marina Gourtovaia
  - & Theo Sanderson
- **Funding acquisition, and leadership and supervision**
  - Christophe Fraser
  - , Jonathan Edgeworth
  - , Judith Breuer
  - , Stephen L. Michell
  - & John A. Todd
- **Funding acquisition and project administration**
  - Michaela John
  - & David Buck
- **Leadership and supervision, and metadata curation**

- Kavitha Gajee
- & Gemma L. Kay
- **Leadership and supervision, and project administration**
  - Sharon J. Peacock
  - & David Heyburn
- **Leadership and supervision, and samples and logistics**
  - Katie Kitchman
  - , Alan McNally
  - , David T. Pritchard
  - , Samir Dervisevic
  - , Peter Muir
  - , Esther Robinson
  - , Barry B. Vipond
  - , Newara A. Ramadan
  - , Christopher Jeanes
  - , Danni Weldon
  - , Jana Catalan
  - & Neil Jones
- **Leadership and supervision, and sequencing and analysis**
  - Ana da Silva Filipe
  - , Chris Williams
  - , Marc Fuchs
  - , Julia Miskelly
  - , Aaron R. Jeffries
  - , Karen Oliver
  - & Naomi R. Park
- **Metadata curation, and samples and logistics**
  - Amy Ash

- , Cherian Koshy
- , Magdalena Barrow
- , Sarah L. Buchan
- , Anna Mantzouratou
- , Gemma Clark
- , Christopher W. Holmes
- , Sharon Campbell
- , Thomas Davis
- , Ngee Keong Tan
- , Julianne R. Brown
- , Kathryn A. Harris
- , Stephen P. Kidd
- , Paul R. Grant
- , Li Xu-McCrae
- , Alison Cox
- , Pinglawathee Madona
- , Marcus Pond
- , Paul A. Randell
- , Karen T. Withell
- , Cheryl Williams
- , Clive Graham
- , Rebecca Denton-Smith
- , Emma Swindells
- , Robyn Turnbull
- , Tim J. Sloan
- , Andrew Bosworth
- , Stephanie Hutchings
- , Hannah M. Pymont
- , Anna Casey
- , Liz Ratcliffe
- , Christopher R. Jones
- , Bridget A. Knight
- , Tanzina Haque
- , Jennifer Hart
- , Dianne Irish-Tavares
- , Eric Witele
- , Craig Mower

- , Louisa K. Watson
- , Jennifer Collins
- , Gary Eltringham
- , Dorian Crudgington
- , Ben Macklin
- , Miren Ituriza-Gomara
- , Anita O. Lucaci
- & Patrick C. McClure

- **Metadata curation, and sequencing and analysis**

- Matthew Carlile
- , Nadine Holmes
- , Christopher Moore
- , Nathaniel Storey
- , Stefan Rooke
- , Gonzalo Yebra
- , Noel Craine
- , Malorie Perry
- , Nabil-Fareed Alikhan
- , Stephen Bridgett
- , Kate F. Cook
- , Christopher Fearn
- , Salman Goudarzi
- , Ronan A. Lyons
- , Thomas Williams
- , Sam T. Haldenby
- , Jillian Durham
- & Steven Leonard

- **Metadata curation, and software and analysis tools**

- Robert M. Davies

- **Project administration, and samples and logistics**

- Rahul Batra
- , Beth Blane
- , Moira J. Spyer
- , Perminder Smith
- , Mehmet Yavus
- , Rachel J. Williams
- , Adhyana I. K. Mahanama
- , Buddhini Samaraweera
- , Sophia T. Girgis
- , Samantha E. Hansford
- , Angie Green
- , Charlotte Beaver
- , Katherine L. Bellis
- , Matthew J. Dorman
- , Sally Kay
- , Liam Prestwood
- & Shavanti Rajatileka

- **Project administration, and sequencing and analysis**
  - Joshua Quick
- **Project administration, and software and analysis tools**
  - Radoslaw Poplawski
- **Samples and logistics, and sequencing and analysis**
  - Nicola Reynolds
  - , Andrew Mack
  - , Arthur Morriss
  - , Thomas Whalley
  - , Bindi Patel
  - , Iliana Georgana
  - , Myra Hosmillo
  - , Malte L. Pinckert

- , Joanne Stockton
- , John H. Henderson
- , Amy Hollis
- , William Stanley
- , Wen C. Yew
- , Richard Myers
- , Alicia Thornton
- , Alexander Adams
- , Tara Annett
- , Hibo Asad
- , Alec Birchley
- , Jason Coombes
- , Johnathan M. Evans
- , Laia Fina
- , Bree Gatica-Wilcox
- , Lauren Gilbert
- , Lee Graham
- , Jessica Hey
- , Ember Hilvers
- , Sophie Jones
- , Hannah Jones
- , Sara Kumziene-Summerhayes
- , Caoimhe McKerr
- , Jessica Powell
- , Georgia Pugh
- , Sarah Taylor
- , Alexander J. Trotter
- , Charlotte A. Williams
- , Leanne M. Kermack
- , Benjamin H. Foulkes
- , Marta Gallis
- , Hailey R. Hornsby
- , Stavroula F. Louka
- , Manoj Pohare
- , Paige Wolverson
- , Peijun Zhang
- , George MacIntyre-Cockett

- , Amy Trebes
- , Robin J. Moll
- , Lynne Ferguson
- , Emily J. Goldstein
- , Alasdair Maclean
- & Rachael Tomb

- **Samples and logistics, and software and analysis tools**
  - Igor Starinskij
- **Sequencing and analysis, and software and analysis tools**
  - Laura Thomson
  - , Joel Southgate
  - , Moritz U. G. Kraemer
  - , Jayna Raghwani
  - , Alex E. Zarebski
  - , Olivia Boyd
  - , Lily Geidelberg
  - , Chris J. Illingworth
  - , Chris Jackson
  - , David Pascall
  - , Sreenu Vattipally
  - , Timothy M. Freeman
  - , Sharon N. Hsu
  - , Benjamin B. Lindsey
  - , Keith James
  - , Kevin Lewis
  - , Gerry Tonkin-Hill
  - & Jaime M. Tovar-Corona
- **Sequencing and analysis, and visualization**
  - MacGregor Cox

- **Software and analysis tools, and visualization**

- Khalil Abudahab
- , Mirko Menegazzo
- , Ben E. W. Taylor
- , Corin A. Yeats
- , Afrida Mukaddas
- , Derek W. Wright
- , Leonardo de Oliveira Martins
- , Rachel Colquhoun
- , Verity Hill
- , Ben Jackson
- , J. T. McCrone
- , Nathan Medd
- , Emily Scher
- & Jon-Paul Keatley

- **Leadership and supervision**

- Tanya Curran
- , Sian Morgan
- , Patrick Maxwell
- , Ken Smith
- , Sahar Eldirdiri
- , Anita Kenyon
- , Alison H. Holmes
- , James R. Price
- , Tim Wyatt
- , Alison E. Mather
- , Timofey Skvortsov
- & John A. Hartley

- **Metadata curation**

- Martyn Guest
- , Christine Kitchen

- , Ian Merrick
- , Robert Munn
- , Beatrice Bertolusso
- , Jessica Lynch
- , Gabrielle Vernet
- , Stuart Kirk
- , Elizabeth Wastnedge
- , Rachael Stanley
- , Giles Idle
- , Declan T. Bradley
- , Jennifer Poyner
- & Matilde Mori

- **Project administration**

- Owen Jones
- , Victoria Wright
- , Ellena Brooks
- , Carol M. Churcher
- , Mireille Fragakis
- , Katerina Galai
- , Andrew Jermy
- , Sarah Judges
- , Georgina M. McManus
- , Kim S. Smith
- , Elaine Westwick
- , Stephen W. Attwood
- , Frances Bolt
- , Alisha Davies
- , Elen De Lacy
- , Fatima Downing
- , Sue Edwards
- , Lizzie Meadows
- , Sarah Jeremiah
- , Nikki Smith
- & Luke Foulser

- **Samples and logistics**
  - Themoula Charalampous
  - , Amita Patel
  - , Louise Berry
  - , Tim Boswell
  - , Vicki M. Fleming
  - , Hannah C. Howson-Wells
  - , Amelia Joseph
  - , Manjinder Khakh
  - , Michelle M. Lister
  - , Paul W. Bird
  - , Karlie Fallon
  - , Thomas Helmer
  - , Claire L. McMurray
  - , Mina Odedra
  - , Jessica Shaw
  - , Julian W. Tang
  - , Nicholas J. Willford
  - , Victoria Blakey
  - , Veena Raviprakash
  - , Nicola Sheriff
  - , Lesley-Anne Williams
  - , Theresa Feltwell
  - , Luke Bedford
  - , James S. Cargill
  - , Warwick Hughes
  - , Jonathan Moore
  - , Susanne Stonehouse
  - , Laura Atkinson
  - , Jack C. D. Lee
  - , Divya Shah
  - , Adela Alcolea
  - , Natasha Ohemeng-Kumi
  - , John Ramble
  - , Jasveen Sehmi
  - , Rebecca Williams

- , Wendy Chatterton
- , Monika Pusok
- , William Everson
- , Anibolina Castigador
- , Emily Macnaughton
- , Kate El Bouzidi
- , Temi Lampejo
- , Malur Sudhanva
- , Cassie Breen
- , Graciela Sluga
- , Shazaad S. Y. Ahmad
- , Ryan P. George
- , Nicholas W. Machin
- , Debbie Binns
- , Victoria James
- , Rachel Blacow
- , Lindsay Coupland
- , Louise Smith
- , Edward Barton
- , Debra Padgett
- , Garren Scott
- , Aidan Cross
- , Mariyam Mirfenderesky
- , Jane Greenaway
- , Kevin Cole
- , Phillip Clarke
- , Nichola Duckworth
- , Sarah Walsh
- , Kelly Bicknell
- , Robert Impey
- , Sarah Wyllie
- , Richard Hopes
- , Chloe Bishop
- , Vicki Chalker
- , Ian Harrison
- , Laura Gifford
- , Zoltan Molnar

- , Cressida Auckland
- , Cariad Evans
- , Kate Johnson
- , David G. Partridge
- , Mohammad Raza
- , Paul Baker
- , Stephen Bonner
- , Sarah Essex
- , Leanne J. Murray
- , Andrew I. Lawton
- , Shirelle Burton-Fanning
- , Brendan A. I. Payne
- , Sheila Waugh
- , Andrea N. Gomes
- , Maimuna Kimuli
- , Darren R. Murray
- , Paula Ashfield
- , Donald Dobie
- , Fiona Ashford
- , Angus Best
- , Liam Crawford
- , Nicola Cumley
- , Megan Mayhew
- , Oliver Megram
- , Jeremy Mirza
- , Emma Moles-Garcia
- , Benita Percival
- , Leah Ensell
- , Helen L. Lowe
- , Laurentiu Maftei
- , Matteo Mondani
- , Nicola J. Chaloner
- , Benjamin J. Cogger
- , Lisa J. Easton
- , Hannah Huckson
- , Jonathan Lewis
- , Sarah Lowdon

- , Cassandra S. Malone
- , Florence Munemo
- , Manasa Mutingwende
- , Roberto Nicodemi
- , Olga Podplomyk
- , Thomas Somassa
- , Andrew Beggs
- , Alex Richter
- , Claire Cormie
- , Joana Dias
- , Sally Forrest
- , Ellen E. Higginson
- , Mailis Maes
- , Jamie Young
- , Rose K. Davidson
- , Kathryn A. Jackson
- , Lance Turtle
- , Alexander J. Keeley
- , Jonathan Ball
- , Timothy Byaruhanga
- , Joseph G. Chappell
- , Jayasree Dey
- , Jack D. Hill
- , Emily J. Park
- , Arezou Fanaie
- , Rachel A. Hilson
- , Geraldine Yaze
- & Stephanie Lo

- **Sequencing and analysis**

- Safiah Afifi
- , Robert Beer
- , Joshua Maksimovic
- , Kathryn McCluggage Masters
- , Karla Spellman
- , Catherine Bresner

- , William Fuller
- , Angela Marchbank
- , Trudy Workman
- , Ekaterina Shelest
- , Johnny Debebe
- , Fei Sang
- , Marina Escalera Zamudio
- , Sarah Francois
- , Bernardo Gutierrez
- , Tetyana I. Vasylyeva
- , Flavia Flaviani
- , Manon Ragonnet-Cronin
- , Katherine L. Smollett
- , Alice Broos
- , Daniel Mair
- , Jenna Nichols
- , Kyriaki Nomikou
- , Lily Tong
- , Ioulia Tsatsani
- , Sarah O'Brien
- , Steven Rushton
- , Roy Sanderson
- , Jon Perkins
- , Seb Cotton
- , Abbie Gallagher
- , Elias Allara
- , Clare Pearson
- , David Bibby
- , Gavin Dabrera
- , Nicholas Ellaby
- , Eileen Gallagher
- , Jonathan Hubb
- , Angie Lackenby
- , David Lee
- , Nikos Manesis
- , Tamyo Mbisa
- , Steven Platt

- , Katherine A. Twohig
- , Mari Morgan
- , Alp Aydin
- , David J. Baker
- , Ebenezer Foster-Nyarko
- , Sophie J. Prosolek
- , Steven Rudder
- , Chris Baxter
- , Sílvia F. Carvalho
- , Deborah Lavin
- , Arun Mariappan
- , Clara Radulescu
- , Aditi Singh
- , Miao Tang
- , Helen Morcrette
- , Nadua Bayzid
- , Marius Cotic
- , Carlos E. Balcazar
- , Michael D. Gallagher
- , Daniel Maloney
- , Thomas D. Stanton
- , Kathleen A. Williamson
- , Robin Manley
- , Michelle L. Michelsen
- , Christine M. Sambles
- , David J. Studholme
- , Joanna Warwick-Dugdale
- , Richard Eccles
- , Matthew Gemmell
- , Richard Gregory
- , Margaret Hughes
- , Charlotte Nelson
- , Lucille Rainbow
- , Edith E. Vamos
- , Hermione J. Webster
- , Mark Whitehead
- , Claudia Wierzbicki

- , Adrienn Angyal
- , Luke R. Green
- , Max Whiteley
- , Emma Betteridge
- , Iraad F. Bronner
- , Ben W. Farr
- , Scott Goodwin
- , Stefanie V. Lensing
- , Shane A. McCarthy
- , Michael A. Quail
- , Diana Rajan
- , Nicholas M. Redshaw
- , Carol Scott
- , Lesley Shirley
- & Scott A. J. Thurston

- **Software and analysis tools**

- Will Rowe
- , Amy Gaskin
- , Thanh Le-Viet
- , James Bonfield
- , Jennifer Liddle
- & Andrew Whitwham

## **Corresponding authors**

Correspondence to [Jeffrey C. Barrett](#) or [Moritz Gerstung](#).

## **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

Vöhringer, H.S., Sanderson, T., Sinnott, M. *et al.* Publisher Correction: Genomic reconstruction of the SARS CoV-2 epidemic in England. *Nature* **606**, E18 (2022). <https://doi.org/10.1038/s41586-022-04887-8>

- Published: 07 June 2022
- Issue Date: 23 June 2022
- DOI: <https://doi.org/10.1038/s41586-022-04887-8>

### Share this article

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

[Get shareable link](#)

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

[Copy to clipboard](#)

Provided by the Springer Nature SharedIt content-sharing initiative

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

# Nature Outlook

- **[Smell](#)** [ 22 Jun 2022 ]
  - Our sense of smell has a reputation for being underdeveloped — but this is ill-founded. The events of the past two years, in which millions of people lost their sense of smell to COVID-19, have put the spotlight on research emerging from the world of olfaction.



## **The science of smell steps into the spotlight**

With millions of people losing their ability to detect aromas as a result of COVID-19, our most underappreciated sense is drawing researchers' attention.

- Richard Hodson

Outlook 22 Jun 2022 [Nature](#)



## **Unpicking the link between smell and memories**

The ability of aromas to bring back highly specific memories is becoming better understood, and could be used to boost and heal our brains.

- Roxanne Khamsi

Outlook 22 Jun 2022 [Nature](#)



## The science behind COVID's assault on smell

The loss of the sense of smell has been a hallmark symptom of COVID-19. The mechanisms behind SARS-CoV-2's ability to interfere with this sense — as well as why variants such as Omicron do so less frequently — are becoming clearer.

- Elie Dolgin

Outlook 22 Jun 2022 [Nature](#)



## How to bring back the sense of smell

Treatments for olfactory loss are currently scarce, but with millions of people unable to smell as a result of COVID-19, researchers are pursuing the problem with renewed vigour.

- Sarah DeWeerdt

Outlook 22 Jun 2022 [Nature](#)

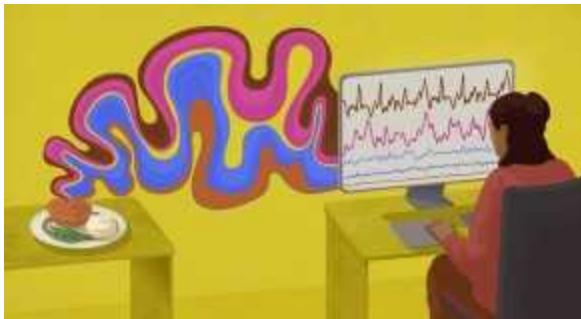


## The dogs learning to sniff out disease

Veterinarian Cynthia Otto explains how we might harness animals' ability to smell human illnesses — including COVID-19.

- Julianna Photopoulos

Outlook 22 Jun 2022 [Nature](#)

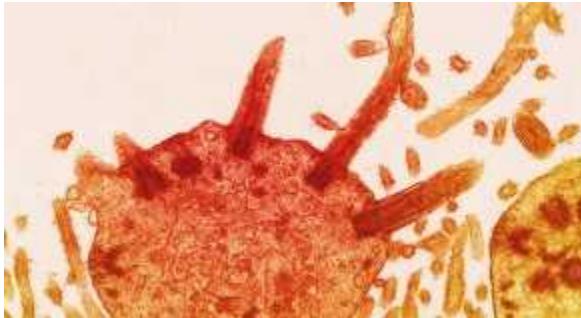


## Restoring smell with an electronic nose

Development of an olfactory implant that could tackle anosmia is in its early stages.

- Simon Makin

Outlook 22 Jun 2022 [Nature](#)



## Olfactory receptors are not unique to the nose

The hundreds of receptors that give us our sense of smell have been found to have important roles in other parts of the body, and the prospect of targeting them with drugs is growing.

- Liam Drew

Outlook 22 Jun 2022 [Nature](#)

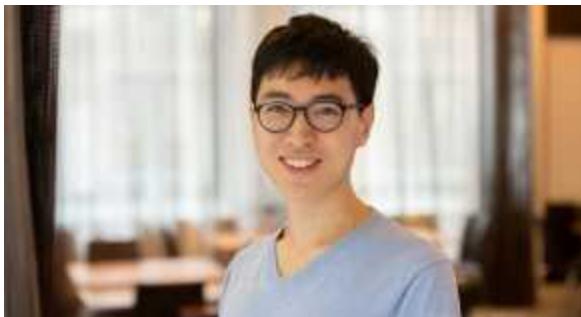


## Sniffing out smell's effects on human behaviour

Olfaction could influence how people respond to threats or select a partner. To investigate, researchers need to design experiments that can capture its effects.

- Michael Eisenstein

Outlook 22 Jun 2022 [Nature](#)



## Building neural networks that smell like a brain

Computational neuroscientist Guangyu Robert Yang lifts the lid on the use of machine learning to detect and process odours, and the wider implications for neuroscience.

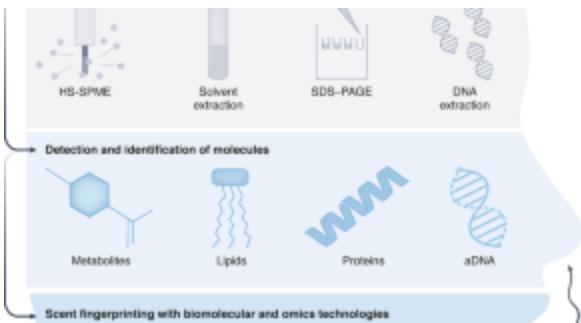
- Conor Purcell

Outlook 22 Jun 2022 [Nature](#)



## From molecules to perception: 126 years at the forefront of olfactory science

Sponsor Feature 22 Jun 2022

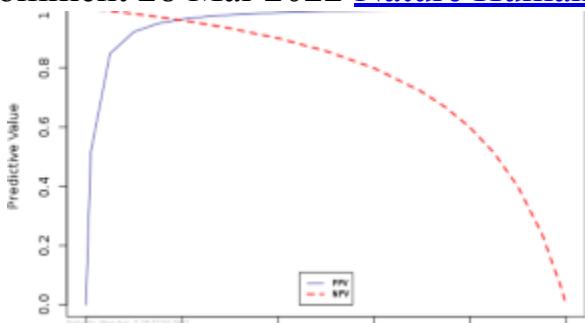


## How to use modern science to reconstruct ancient scents

Olfaction has profoundly shaped human experience and behaviour from the deep past through to the present day. Advanced biomolecular and ‘omics’ sciences enable more direct insights into past scents, offering new options to explore critical aspects of ancient society and lifeways as well as the historical meanings of smell.

- Barbara Huber
- Thomas Larsen
- Nicole Boivin

Comment 28 Mar 2022 [Nature Human Behaviour](#)

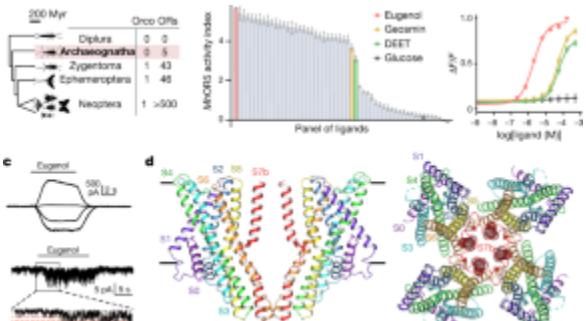


## The detection dogs test is more sensitive than real-time PCR in screening for SARS-CoV-2

Hag-Ali and colleagues highlight the potential for using trained dogs for detecting COVID-19 positive patients. The dogs, originally trained for explosives detection, were able to detect COVID-19 positive sweat samples with a sensitivity rivaling the gold-standard RT-PCR test currently used.

- Mohammed Hag-Ali
- Abdul Salam AlShamsi
- Abdullatif Alteraifi AlShamsi

Article Open Access 3 Jun 2021 [Communications Biology](#)

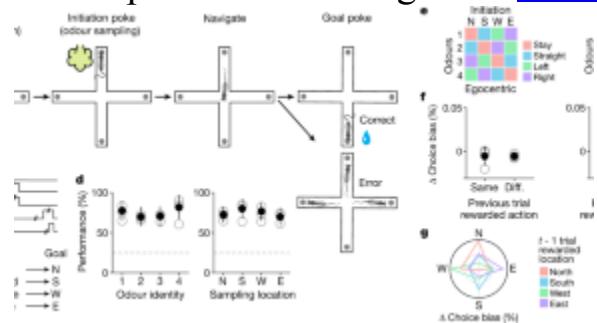


## The structural basis of odorant recognition in insect olfactory receptors

Structural and functional analysis of an insect olfactory receptor shed light on how receptors can be activated by diverse odorants.

- Josefina del Mármol
- Mackenzie A. Yedlin
- Vanessa Ruta

Article Open Access 4 Aug 2021 [Nature](#)

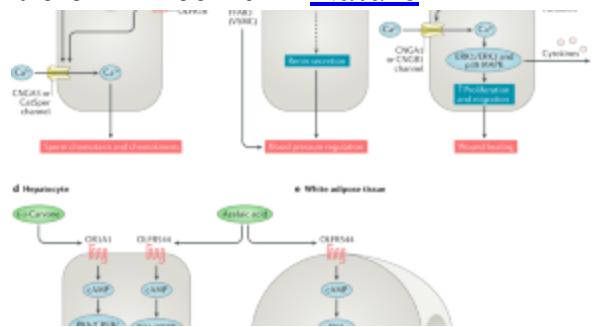


## Spatial maps in piriform cortex during olfactory navigation

Studies using neural ensemble recordings in rats show that cells in the piriform cortex carry a spatial representation of the environment and link locations to olfactory sensory inputs.

- Cindy Poo
- Gautam Agarwal
- Zachary F. Mainen

Article 22 Dec 2021 [Nature](#)

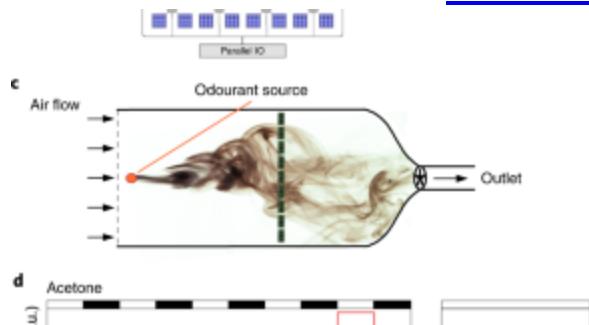


## [Therapeutic potential of ectopic olfactory and taste receptors](#)

Olfactory and taste receptors are ectopically expressed in multiple extra-nasal and extra-oral tissues, exhibiting potential functions in a diverse range of biological processes. Here, Lee et al. discuss the physiological roles of these ectopic olfactory and taste receptors, assessing their emerging therapeutic and diagnostic potential in conditions including asthma, wound healing, obesity and cancer.

- Sung-Joon Lee
- Inge Depoortere
- Hanns Hatt

Review Article 30 Nov 2018 [Nature Reviews Drug Discovery](#)



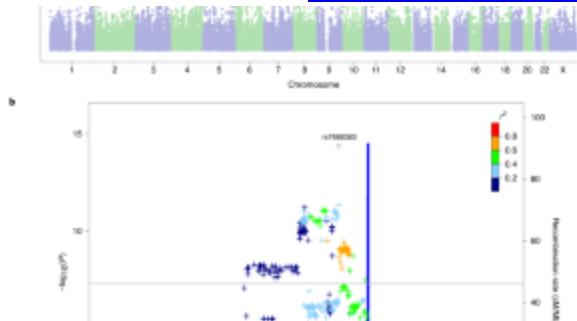
## [Rapid online learning and robust recall in a neuromorphic olfactory circuit](#)

Integrating knowledge about the circuit-level organization of the brain into neuromorphic artificial systems is a challenging research problem. The authors present a neural algorithm for the learning of odourant signals and

their robust identification under noise, based on the architecture of the mammalian olfactory bulb and implemented on the Intel Loihi neuromorphic system.

- Nabil Imam
- Thomas A. Cleland

Article 16 Mar 2020 [Nature Machine Intelligence](#)

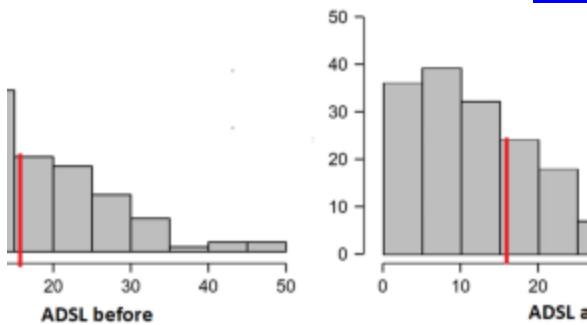


### The *UGT2A1/UGT2A2* locus is associated with COVID-19-related loss of smell or taste

Multi-ancestry genome-wide analyses identify variants near *UGT2A1* and *UGT2A2* associated with COVID-19-related loss of smell or taste. Both genes are expressed in the olfactory epithelium and play a role in metabolizing odorants.

- Janie F. Shelton
- Anjali J. Shastri
- Adam Auton

Brief Communication 17 Jan 2022 [Nature Genetics](#)



## Symptoms of depression change with olfactory function

- Agnieszka Sabiniewicz
- Leonie Hoffmann
- Thomas Hummel

Article Open Access 5 Apr 2022 [Scientific Reports](#)

---

This article was downloaded by **calibre** from  
<https://www.nature.com/collections/hcjfdcbgaj>.

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